Growth Modeling in a Diagnostic Classification Model (DCM) Framework

By

Qianqian Pan
M.A., East China Normal University, 2014
B.A., East China Normal University, 2011

Submitted to the graduate degree program in Educational Psychology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Chair: Neal Martin Kingston

Meagan M Patterson

Jonathan Templin

William P Skorupski

Lesa Hoffman

Date Defended: July 30, 2018

The dissertation committee for Qianqian Pan certifies that this is the approved version of the following dissertation:

Growth Modeling in a Diagnostic Classification Model (DCM) Framework

Chair: Neal Martin Kingston

Date Approved: July 30, 2018

Abstract

A multivariate longitudinal DCM is developed that is the composite of two components, the log-linear cognitive diagnostic model (LCDM) as the measurement model component that evaluates the mastery status of attributes at each measurement occasion, and a generalized multivariate growth curve model that describes the growth of each attribute over time. The proposed model represents an improvement in the current longitudinal DCMs given its ability to incorporate both balanced and unbalanced data and to measure the growth of a single attribute directly without assuming that attributes grow in the same pattern. One simulation study was conducted to evaluate the proposed model in terms of the convergence rates, the accuracy of classification, and parameter recoveries under different combinations of four design factors: the sample size, the growth patterns, the G matrix design, and the number of measurement occasions.

The results revealed the following: (1) In general, the proposed model provided good convergence rates under different conditions. (2) Regarding the classification accuracy, the proposed model achieved good recoveries on the probabilities of attribute mastery. However, the correct classification rates depended on the cutpoint that was used to classify individuals. For individuals who truly mastered the attributes, the correct classification rates increased as the measurement occasions increased; however, for individuals who truly did not master the attributes, the correct classification rates decreased slightly as the numbers of measurement occasions increased. Cohen's kappa increased as the number of measurement occasions increased. (3) Both the intercept and main effect parameters in the LCDM were recovered well. The interaction effect parameters had a relatively large bias under the condition with samll sample size and fewer measurement occasions; however, the recoveries were improved as the sample size and the number of measurement occasions increased. (4) Overall, the proposed model achieved acceptable recoveries on both the fixed and random effects in the generalized growth curve model.

Acknowledgments

There are many people who helped me through my dissertation and my doctoral program, and I sincerely express the depth of my gratitude to each of them.

Many, many thanks to my advisor, Dr. Neal Kingston, for his mentorship and guidance through my whole doctoral program. Thank you for supporting my choice of such a complicated topic and providing limitless support from the beginning to the end. Also, I am extremely grateful that you have given me different opportunities to explore myself along with your constant encouragement.

Many thanks to Dr. Lesa Hoffman, not only for teaching me about the growth modeling but also for walking me through the whole process including the research design, inspecting possible errors, and correcting format mistakes and typos. I am incredibly appreciative of your advice on how to choose an appropriate career. Also, thank you for your hug and encouraging words. Many thanks to Dr. Jonathon Templin, for opening the door of the Diagnostic Classification Models for me and helping me revise the code to make the simulation more efficient. Without your help, the simulation would not have been finished in time. Thank you to Dr. William Skorupski, for introducing me to Bayesian statistics and simulation methods . Without this knowledge, I would not been able to even start this dissertation. Also, I would like to thank Dr. Meagan Patterson for being very supportive at the defense meeting, and Dr. Bruce Frey for the suggestions at the proposal meeting.

I could not complete the dissertation without many other people's generous help. Thank you to Dr. Jake Thompson for introducing me to Rstudio and teaching me how to use Amazon AWS to complete my simulation as quickly as possible. Thank you to my supervisors, Dr. Brooke Nash and Dr. Amy Clark, being so understanding and supportive throughout my GRA career, especially during the last year when I was working on the dissertation. Many thanks to Dr. Ming Gong, for teaching me how to use Batch to implement parallel processing, providing me computing resources, as well as being a super good peer mentor to encourage me to overcome all difficulties throughout my doctoral program.

Moreover, I would like to express my sincere appreciation to my friends. Thank you to my friends in the states, Xiaolin, Qingqing, Frank, Lu, Pearl, Ge, Zhehan, Jihong, Kim, Shengnan, and Yurou, for being supportive to help me overcome difficulties in my study as well as making my doctoral program as fun as it could be. Also, thank you to my friends in China, Ning, Peipei, Yinxiu, Jing, Shan, Qi, and Lingjie, for always being there for me, even we are thousands of miles apart. And, thank you to my master and bachelor advisors, Dr. Fuyi Yang, and Prof. Fujuan Zhang, for their consistent consideration and continuous encouragement.

Finally, thank you to my parents, Yi and Xiaoxin, for being such supportive, enlightened, and optimism parents, and for their unconditional love, understanding, and encouragement all the time.

Table of Contents

Chapter 1: Introduction ... 11

Background ... 11

Statement of Problem ... 11

Purpose of Study ... 12

Chapter 2 : Literature Review ... 13

Diagnostic Classification Models ... 13

Conceptual Foundations of DCMs. ... 13

Statistical Foundations of DCMs. ... 15

Longitudinal DCMs ... 20

LTA-based longitudinal DCMs. ... 20

Higher-order DCM-based Longitudinal DCMs. ... 23

Chapter 3 : Research Design and Methods ... 28

Proposed Multivariate Longitudinal DCM ... 28

Research questions. ... 29

Simulation Study Design ... 30

Design factors. ... 30

Fixed conditions. ... 33

Data generation procedures. ... 36

Analysis Plan and Outcome Variables ... 37

Evaluation Criteria ... 38

Convergence rates. ... 38

Classification accuracy. ... 39

Parameter recovery. ... 40

Chapter 4 : Results ... 41

Convergence Rates ... 41

Classification Accuracy ... 41

Parameter Recovery ... 51

Measurement model parameter recovery. ... 51

Structural model parameter recovery. ... 56

Chapter 5 : Discussion ... 70

Performance of the Multivariate Longitudinal DCM ... 71

Model convergence. ... 71

Classification accuracy. ... 71

Parameter recoveries. 72
Conclusion and Recommendations 73
Contributions and Limitations 75
Reference 76
Appendix A 79
Appendix B 82

List of Figures

Figure 3.1 Two patterns of growth across attributes 31

List of Tables

Table 2.1: The Hypothetical Q-matrix 14
Table 2.2: The Hypothetical Transition Probability Matrix 21
Table 3.1 G Matrix Specification and Corresponding Correlation Matrix 32
Table 3.2: Q-matrix Design 34
Table 3.3: Initial Level and Growth Rates of Linear Predictors 35
Table 3.4: Base-rates of Attributes over Time 36
Table 4.1: Average Convergence Rates 41
Table 4.2 Bias of Probability of Attribute Mastery (MO=3) 42
Table 4.3 Bias of Probability of Attribute Mastery (MO=5) 43
Table 4.4 Average Correct Classification Rates for Individuals Who Truly Mastered Attribute (MO=3) 45
Table 4.5 Average Correct Classification Rates for Individuals Who Truly did not Master Attribute (MO=3) 45
Table 4.6 Average Correct Classification Rates for Individuals Who Truly Mastered Attribute (MO=5) 46
Table 4.7 Average Correct Classification Rates for Individuals Who Truly did not Mastered Attribute (MO=5) 47
Table 4.8 Average Kappa (MO=3) 49
Table 4.9 Average Kappa (MO=5) 50
Table 4.10 Summary of Measurement Model Parameter Recoveries 52
Table 4.11 ANOVA Results of Measurement Model Parameter Recoveries (MO=3) 54
Table 4.12 ANOVA Results of Measurement Model Parameter Recoveries (MO=5) 55
Table 4.13 Summary of Fixed Effects Recoveries (MO=3) 58
Table 4.14 Summary of Fixed Effects Recoveries (MO=5) 59
Table 4.15 ANOVA Results of Fixed Effects Parameter Recoveries (MO=3) 60
Table 4.16 ANOVA Results of Fixed Effects Parameter Recoveries (MO=5) 61
Table 4.17 Summary of Random Variance Recoveries (MO=3) 63
Table 4.18 Summary of Random Variance Recoveries (MO=5) 64
Table 4.19 ANOVA Results of Random Variance Parameter Recoveries (MO=3) 65
Table 4.20 ANOVA Results of Random Variance Parameter Recoveries (MO=5) 66
Table 4.21 Summary of Covariance Parameter Recoveries 68
Table 4.22 ANOVA Results of Random Covariance Parameter Recoveries (MO=3) 69
Table 4.23 ANOVA Results of Random Covariance Parameter Recoveries (MO=5) 69

List of Equations

Equation 2.1 16
Equation 2.2 17
Equation 2.3 17
Equation 2.4 18
Equation 2.5 19
Equation 2.6 19
Equation 2.7 22
Equation 2.8 22
Equation 2.9 22
Equation 2.10 24
Equation 2.11 25
Equation 2.12 25
Equation 2.13 25
Equation 2.14 25
Equation 2.15 25
Equation 2.16 25
Equation 3.1 29
Equation 3.2 29
Equation 3.3 29
Equation 3.4 29
Equation 3.5 29
Equation 3.6 29
Equation 3.7 37
Equation 3.8 38
Equation 3.9 39
Equation 3.10 40
Equation 3.11 40

Chapter 1: Introduction

Background

Diagnostic classification models (DCMs; e.g., Rupp, Templin, & Henson, 2010), also referred to as cognitive diagnosis models (CDMs; e.g., Leighton & Gierl, 2007), are defined as a family of confirmatory multidimensional latent-variable models with categorical latent variables (Rupp et al., 2010). DCMs evaluate the student's mastery status on each latent variable from a set of narrowly defined latent variables, referred to attributes in the DCM literature, and then classify students into attribute profiles that were determined as a priori (DiBello, Stout, & Roussos, 1995). DCMs provide fine-grained and multidimensional diagnostic information, which could help educators adjust classroom instruction and improve student learning. Since the traditional scale scores (e.g., IRT scores) have limits in providing enough information to inform classroom instruction and learning (e.g., de La Torre, 2009), DCMs have received growing attention in the educational measurement community as well as from educational practitioners in recent years.

Statement of Problem

DCMs have been increasingly used for empirical data analysis in recent years. For example, DCMs have been retrofitted to existing large-scale assessments to identify examinees' mastery status of tested skills (e.g., George & Robitzsch, 2014; Lee & Sawaki, 2009; Ravand, 2016; Sedat & Arican, 2015). In addition, some researchers successfully demonstrated the practical uses of DCMs in test development (Bradshaw, Izsák, Templin, & Jacobson, 2014). DCMs have also been applied in one large-scale assessment program (Dynamic Learning Maps® alternate assessment; DLM®) to detect distinct patterns of skill mastery for students with significant cognitive disabilities. Most applications of DCMs are static -- DCMs are used to classify individuals at a single time point. When longitudinal data are modeled, the longitudinal DCM is used to measure the change in the attribute profiles and mastery status over time.

Currently, two types of longitudinal DCMs have been proposed to analyze longitudinal data in the DCM framework. Latent transition analysis (LTA; Collins & Wugalter, 1992) -based longitudinal DCMs (e.g., Kaya & Leite, 2017; Li, Cohen, Bottge, & Templin, 2016; Madison, 2016) estimates the probabilities of transitioning from one latent class to another latent class or staying at the same latent class across two measurement occasions. Higher-order DCM (HDCM; e.g., de la Torre & Douglas, 2004; Templin & Bradshaw, 2014) -based longitudinal DCMs (e.g., Huang, 2017; Zhan, Jiao, & Liao, 2017) assumes a higher-order continuous factor to predict the mastery status of lower-order attributes so that the changes in the higher-order factor are used to infer the changes of lower-order attributes over time.

These two longitudinal DCM approaches have been evaluated by a few simulation studies and some applied research, which has demonstrated their utility for analyzing longitudinal data in the DCM framework. However, these models are not without limitations. For example, LTA-based longitudinal DCMs are restricted to the balanced data and assume attributes are independent. In addition, LTA-based approach is limited to assessing changes between only two measurement occasions (Huang, 2017).

On the other hand, HDCM-based longitudinal DCMs assume all attributes have similar growth trajectories. However, previous studies found attributes could change in different ways (e.g., Li et al., 2016; Madison, 2016).

Purpose of Study

The overarching goal of the current study is to develop a multivariate longitudinal DCM, improves upon current longitudinal DCMs by (1) being able to incorporate both balanced data and unbalanced data and (2) measuring the growth of multiple attributes that have dissimilar growth trajectories. More specific research questions are presented in Chapter 3.

Chapter 2 : Literature Review

Diagnostic Classification Models

Conceptual Foundations of DCMs.

The definition of DCMs. DCMs have been defined as confirmatory multidimensional latent-variable models by Rupp et al. (2010), and they share some similarities with factor analysis (FA) models, multidimensional item response theory (MIRT) models, and latent class models (LCMs). However, DCMs can be distinguished from these models in several main features, such as (1) the grain size of the constructs; DCMs have more narrowly defined constructs, (2) the nature of the latent variables; DCMs have categorical latent variables rather than continuous, (3) the loading structure; DCMs allow items to load on one or multiple latent variables, and (4) the number of latent classes; DCMs have a pre-determined number of latent classes.

Some terms commonly used in the DCM framework are attributes to describe the discrete latent variables, and Q-matrix, which defines the loading structure. In what follows, more information is provided regarding the characteristics of DCMs and the similarities and differences between DCMs and other conventional latent trait models are further described. A hypothetical mathematics test with 10 items that measure 2 attributes (addition and subtraction) will be used as an example to facilitate the introduction.

The properties of DCMs.

The nature of the constructs. Similar to FA models and MIRT models, DCMs utilize multiple latent attributes to measure individuals. Therefore, DCMs are suitable for modeling multidimensional constructs. However, the difference between DCMs and FA or MIRT is the grain size of the construct. DCMs can be used to measure more narrowly defined constructs than traditional multidimensional models. Consequently, DCMs have the potential to provide fine-grained diagnostic information (Rupp et al., 2010, pp. 83-84).

The nature of the latent variables. The latent variables in DCMs are discrete, which is similar to the Latent Class Analysis (LCA; e.g., Goodman, 1974; Lazarsfeld & Henry, 1968), meaning that individuals are classified into different latent classes. In the current study, only dichotomous latent variables are included. However, polytomous latent variables could be incorporated in the DCMs as well (e.g., Chen & Torre, 2013; Davier, 2005).

The nature of the model structure. DCMs are confirmatory in the model structure because the loading structure, often called the Q-matrix, is specified a priori and the fit of the model can be assessed after data are collected. Table 2.1 presents a hypothetical Q-matrix of the example, where rows represent items and columns represent attributes. An entry of 1 means this item measures one particular attribute, an entry of 0 indicates it does not measure this attribute. Generally, LCA allows items to load on all latent variables, and it does not specify the loading structure prior to the analysis. However, the use of a Q-matrix will impact the calibration process and the interpretations of the model results. Misspecification of the Q-matrix will lead to inaccurate classification results (e.g., de La Torre, 2008; de La Torre, Hong, & Deng, 2010; Rupp & Templin, 2008a).

Table 2.1

The Hypothetical Q-matrix

	A1	A2
I1	1	0
I2	1	0
I3	1	0
I4	1	0
I5	1	1
I6	0	1
I7	0	1
I8	0	1
I9	0	1
I10	1	1

Note. A1 refers to Attribute 1. A2 refers to Attribute 2.

The complexity of the Q-matrix. The Q-matrix can indicate a simple relationship between items and attributes (e.g., each item measures one and only one attribute) or a complex relationship (e.g., all or some items might measure one or more attributes). Following the previous example shown in Table 2.1, I1 and I5 are as follows:

I1: 5 + 4 = $\square$

I5: 5 + 3 – 2 = $\square$

Obviously, I1 only measures the first attribute, addition, and I5 measures both attributes, addition and subtraction, reflecting within-item multidimensionality. Consequently, this complex loading structure allows for interaction effects between attributes (Kunina-Habenicht, Rupp, & Wilhelm, 2012). However, items commonly only measure one latent variable in MIRT or FA models, reflecting between-item multidimensionality only.

The number of latent classes. DCMs aim to estimate individual mastery status for each attribute and, then, classify each individual into one latent class with an identical attribute profile. The number of possible latent classes is determined prior to the analysis, which is decided by the total number of attributes and their dependencies (Rupp et al., 2010, pp. 87-88). The total number of possible latent classes for A independent dichotomous attributes is 2^{A}, and the total number of distinct items that can measure these attributes is $2^{\mathrm{A}-1}$. In the aforementioned example, two dichotomous attributes are measured; thus, there are up to 4 latent classes, which are often called latent profiles in the DCM literature. The possible latent profiles are listed as follows: $\mathrm{A}_1 = \{0,0\}, \mathrm{A}_2 = \{0,1\}, \mathrm{A}_3 = \{1,0\}, \mathrm{A}_4 = \{1,1\}$, where an entry of 0 represents the nonmastery of one particular attribute, and an entry of 1 represents the mastery status. In addition, the meaning of each latent profile is known a priori. For example, A_1 represents nonmastery for both attributes, and A_2 represents the mastery of Attribute 2 but the nonmastery of Attribute 1.

Statistical Foundations of DCMs.

The overall objective of DCMs is to determine mastery or nonmastery of attributes of individuals and classify individuals into different latent profiles. To achieve this objective, a DCM is composite of two models, the measurement model and the structural model. The measurement model estimates the probability of observing each item response given the latent profile membership, namely, the item response probability, and the structural model estimates the probability of latent profile membership, namely, the latent profile (or latent class) membership probabilities (Rupp et al., 2010, p. 114).

Structural model. Suppose that there is a total of C latent classes. Then, in the structural component, the latent class membership probability ν_c is estimated. Each individual is a member of one and only one latent class, indicating the latent classes are mutually exclusive and exhaustive. Moreover, all class membership probabilities sum up to 1 (See Equation 2.1). Therefore, a total of $\mathrm{C} - 1$ class membership probabilities need to be estimated.

$$\sum_{c=1}^{C} \nu_c = 1 \qquad \text{Equation 2.1}$$

As the structural model in the DCMs, ν_c, also represents the base-rate probability of the latent class membership in the population. Similar to the role of the structural model in the structural equation model (SEM; e.g., Brown, 2014) that describes the relationships among latent variables, the structural model in the DCMs also reflects the relationships among attributes.

Higher-order DCMs. With the increasing of the numbers of attributes, the complexity of the structural model is increased quickly. Therefore, higher-order DCMs, also referred to hierarchical DCMs (HDCM;e.g., de la Torre & Douglas, 2004; Rupp & Templin, 2008b), are developed to constrain structural parameters in some certain way to result in a more parsimonious structural model (Rupp et al., 2010, pp. 169-190). Several approaches have been applied to parameterize HDCMs, including the log-linear models (Rupp & Templin, 2008b), tetrachoric correlation models (Hartz, 2002), logistic regression models (de la Torre & Douglas, 2004), and probit models (Templin, Henson, Templin, & Roussos, 2008). For example, Rupp and Templin (2008b) adopted a log-linear model to predict $\boldsymbol{\nu_c}$ through the main effects of attributes and interaction effects for all possible combinations of attributes, which allows us to constrain or remove the interaction parameters to reduce the complexity of the structural model. Moreover, Hartz (2002) utilized the tetrachoric correlations between pairs of attributes to describe the relationships among attributes, such that fewer numbers of higher-order factors could be fit into the structural model to account for the relations among lower-order attributes, resulting in less structural parameters. Furthermore, some researchers have adopted the IRT framework, where the probabilities of mastering attributes are predicted by a higher-order continuous latent variable though some link functions, including the logistic model and the probit model (e.g., de la Torre & Douglas, 2004; Templin et al., 2008).

Measurement model. The measurement model of DCMs is described in Equation 2.2, which expresses the probability of a correct item response (π_{ic}) to item i by individual r conditional on the latent class c.

$$P(X_{ri} = 1|c) = \pi_{ic} \qquad \text{Equation 2.2}$$

The current study focuses only on the dichotomous item response, so the correct response is defined as 1. However, the measurement model allows the polytomous and continuous item responses as well (Rupp & Templin, 2008b).

Finally, the general equation of the probabilities in the DCMs is described in Equation 2.3, which expresses the probability of observing a vector of item response X_r is a function of the probability of a correct item response, π_{ic}, and the probability of being in the latent class c.

$$P(X_r = x_r) = \sum_{c=1}^{C} \nu_c \prod_{i=1}^{I} \pi_{ic}^{x_{ir}} (1 - \pi_{ic})^{1-x_{ir}} \qquad \text{Equation 2.3}$$

Different types of measurement models. Various types of measurement models have been proposed to parameterize the item response probability π_{ic} in the DCM literature, including the *deterministic-input, noisy-and-gate* (DINA) model (e.g., Junker & Sijtsma, 2001; Torre, 2009), the *noisy-input, deterministic-*

and-gate (NIDA) model (e.g.,Junker & Sijtsma, 2001; Maris, 1999), the *non-compensatory re-parameterized unified* (NC-RUM) model (Roussos et al., 2007), the *deterministic input, noisy-or-gate* (DINO) model (Templin & Henson, 2006), the *noisy input, deterministic-or-gate* (NIDO) model (Rupp et al., 2010), the *compensatory re-parameterized unified* (C-RUM) model (Rupp & Templin, 2008b), the generalized DINA model (de la Torre, 2011), the log-linear cognitive diagnostic model (LCDM; Rupp et al., 2010), and the general diagnostic model (GDM; von Davier, 2005). These models differ from each other in how they parameterize the item response probability π_{ic}, for example, some models could predict the observed variables in a non-compensatory way (e.g., DINA, NIDA), while some models could predict the item response in a compensatory way (e.g., DINO, NIDO). The current study employs the LCDM as the measurement model because it is a more general diagnostic model that subsumes the aforementioned DCMs (e.g., DINA, DINO, NC-RUM, etc.) through constraining some item parameters (Rupp et al., 2010, pp. 144-168).

LCDM used as the measurement model. In the LCDM, the item response probability is a function of attributes measured by the item, the item parameters, and the attributes mastery status of the individual.

The LCDM estimates the item response probability in the Log-linear model (Rupp et al., 2010, p. 114). Therefore, the probability of a correct response to item i by an individual r in the latent class c could be expressed as follows:

$$\pi_{ic} = P(X_{ic} = 1|\alpha_c) = \frac{\exp(\lambda_{i,0} + \lambda_i^T h(\alpha_c, q_i))}{1 + \exp(\lambda_{i,0} + \lambda_i^T h(\alpha_c, q_i))} \qquad \text{Equation 2.4}$$

where q_i is the set of Q-matrix entries for item i; the intercept, $\lambda_{i,0}$, represents the logit of a correct response given that all Q-matrix indicated attributes are not possessed by a respondent; the vector λ_i represents a vector of size $(2^A - 1) \times 1$ with the main effect, and $h(\alpha_c, q_i)$ is a vector of size $(2^A - 1) \times 1$ with linear combinations of the α_c and q_i. Rupp et al. (2010, p. 156) defined the sum of the intercept and the combination of the α_c and q_i as the kernel, which is expressed as follows,

$$kernel = \lambda_{i,o} + \lambda_i^T h(\alpha_c, q_i)$$
$$= \lambda_{i,0} + \sum_{\alpha=1}^{A} \lambda_{i,1,(\alpha)} \alpha_{c\alpha} q_{i\alpha} + \sum_{\alpha=1}^{A} \sum_{\alpha'>1} \lambda_{i,2,(\alpha,\alpha')} \alpha_{c\alpha} \alpha_{c\alpha'}, q_{i\alpha} q_{i\alpha'} + \cdots$$ Equation 2.5

where $\lambda_{i,1,(\alpha)}$ represents the main effect of attribute α, $\lambda_{i,2,(\alpha,\alpha')}$ represents the two-way interaction between attribute α and attribute α', where $\alpha \neq \alpha', q_{i\alpha} \neq 1, q_{i\alpha'} \neq 1$. There is an A-way interaction for items measuring all A attributes, if possible.

The previous hypothesized example is adopted here to demonstrate the LCDM. As shown in Table 2.1, I5 measures two attributes, α_1 and α_2, then the probability of a correct response to I5 given an individual r in latent class c is expressed as:

$$P(X_{5c} = 1|\alpha_c) = \frac{\exp(\lambda_{1,0} + \lambda_{1,1,(1)}(\alpha_1) + \lambda_{1,1,(2)}(\alpha_2) + \lambda_{1,2,(1,2)}(\alpha_1 \cdot \alpha_2))}{1 + \exp(\lambda_{1,0} + \lambda_{1,1,(1)}(\alpha_1) + \lambda_{1,1,(2)}(\alpha_2) + \lambda_{1,2,(1,2)}(\alpha_1 \cdot \alpha_2))}$$ Equation 2.6

As shown in Equation 2.6, the probability of a correct response to I5 is a function of the intercept ($\lambda_{1,0}$), the simple main effects of attribute 1 ($\lambda_{1,1,(1)}$) and attribute 2 ($\lambda_{1,1,(2)}$), interaction effects between these two attributes ($\lambda_{1,2,(1,2)}$), and the mastery status of two attributes. The intercept represents the log-odds of a correct answer for individuals who did not master any of the attributes. The simple main effects of attributes represent the increase in log-odds for individuals who have mastered only one of the attributes. Moreover, the interaction represents the change in log-odds for individuals who have mastered both attributes. Since the attributes are all dichotomous, $\alpha_1 = 1$ indicates attribute 1 is mastered, while $\alpha_1 = 0$ indicates attribute 1 is not mastered. As mentioned, as a general diagnostic model, the LCDM is able to subsume other frequently used DCMs. Using the same example, when two main effects are fixed to 0, the DINA model is achieved (Bradshaw & Madison, 2016).

As discussed above, DCMs are superior to FA, MIRT, and LCA models in many aspects, including classifying individuals into finite latent classes with known meaning and providing diagnostic information.

However, DCMs discussed above could only estimate latent profile probabilities at a single time point, which limits their use in tracking student growth in attributes over time and evaluating the effectiveness of the instruction. Therefore, in recent years, a few longitudinal DCMs have been proposed as the extension of the general DCMs, which uses the repeated measures of individuals to analyze the changes in the mastery status of attributes and/or attribute profiles over time. Few empirical studies and simulation studies have been conducted to demonstrate the use of those longitudinal DCMs. In the following section, some longitudinal DCMs are reviewed and compared with each other.

Longitudinal DCMs

Currently, two types of longitudinal DCMs have been developed and applied to measure longitudinal data, including latent transition analysis (LTA; Collins & Wugalter, 1992)-based longitudinal DCMs (e.g., Kaya & Leite, 2017; Li et al., 2016; Madison, 2016), and Higher-order DCM (HDCM; e.g., de la Torre & Douglas, 2004; Templin & Bradshaw, 2014)-based longitudinal DCMs (e.g., Huang, 2017; Zhan et al., 2017). The definitions, model specifications, and limitations of these two types of longitudinal DCMs are briefly reviewed as follows.

LTA-based longitudinal DCMs.

Latent Transition Analysis (LTA). Latent class analysis (LCA; e.g., Goodman, 1974; Lazarsfeld & Henry, 1968) is developed for analyzing categorical latent variables. Latent transition analysis (LTA) is the extension of the general LCA for longitudinal data, which enables the estimation of both the latent class membership probability, often called the latent status prevalence in the LTA, and the probabilities of transitions in latent status from one measurement occasion to the next (Lanza, Flaherty, & Collins, 2003, p. 161).

A total of three sets of parameters are estimated in the LTA, including the latent status prevalence, item response probabilities, and transition probabilities. Latent status prevalence represents the probabilities of belonging to each latent class at each measured occasion. Transition probabilities are estimated as the primary interests in the LTA, which expresses the probabilities of transitioning from one latent status to another status or staying in the same status, conditional on the previous latent status membership (Lanza et al., 2003, p. 196). Suppose there are T measurement occasions; thus, a total of $\mathrm{T}-1$ transition probabilities could be estimated. For example, in the aforementioned example, these two attributes have been repeatedly measured at two measurement occasions, namely, Time 1 and Time 2. Table 2.2 presents the hypothetical transition probability. The diagonal elements represent the probabilities of remaining the same latent status across measurement occasions, conditional on the latent status at Time 1. Additionally, the off-diagonal elements represent the probabilities of transitioning to other latent statuses at Time 2 given a particular latent status at Time 1.

Table 2.2

The Hypothetical Transition Probability Matrix

	Time 2			
Time 1	class1(0,0)	class2(0,1)	class3(1,0)	class4(1,1)
class1(0,0)	.49	.23	.22	.06
class2(0,1)	.06	.55	.09	.29
class3(1,0)	.12	.11	.48	.30
class4(1,1)	.00	.03	.04	.92

Note. Numbers in the parenthesis represents the latent profiles. (0,0) represents nonmastery of two attributes; (0,1) represents the mastery of Attribute 1 and nonmastery of Attribute 2; (1,0) represents the nonmastery of Attribute 1 and mastery of Attribute 2; (1,1) represents the mastery of both Attributes 1 and 2.

LTA-based longitudinal DCMs. LTA-based longitudinal DCMs are a composite of the DCM, as the measurement model to classify individuals into different latent classes at each time point, and the LTA, as the structural model to estimate the transition probability to represent the changes in latent class membership across two measurement occasions.

Suppose there are $\mathrm{i}=1,\ldots,\mathrm{I}$ items, and these items have been tested at $\mathrm{t}=1,\ldots,\mathrm{T}$ times. Thus, the vector of item response for individual r is represented by $X_r=(x_{1,1},\ldots,x_{I,T})$. Let S_1 represent the categorical latent variable at Time 1, where $S_1=1,\ldots,S$; thus, S_T represents the categorical latent variable at time T, where $S_T=1,\ldots,S$. The number of latent statuses is identical across measurement occasions (i.e.,

$S_1 = S_2 = \cdots = S_T = S$). Let ν_{s_t} represent the prevalence of latent status s at time t, meaning the probability of membership in latent status s at time t.

Equation 2.7 describes the probability of observing a vector of a total of I items responses by individual r given the latent status S_t, where π_{iS_t}, the probability of a correct answer to item i given the latent status, is estimated via a DCM.

Equation 2.8 presents the conditional transition probabilities, where $\tau_{S_{t+1}|S_t}$ represents the probability of a transition to latent status s at time $t+1$, conditional on membership in latent status s at time t.

Equation 2.9 presents the general expression of the LTA-based longitudinal DCM.

$$P(X_r = x_r | S_t) = \prod_{i=1}^{I} \pi_{iS_t}^{x_{rt}} \left(1 - \pi_{iS_t}\right)^{1 - x_{rit}} \qquad \text{Equation 2.7}$$

$$\begin{bmatrix} \tau_{1_{t+1}|1_t} & \tau_{2_{t+1}|1_t} & \cdots & \tau_{S_{t+1}|1_t} \\ \tau_{1_{t+1}|2_t} & \tau_{2_{t+1}|2_t} & \cdots & \tau_{S_{t+1}|2_t} \\ \cdots & \cdots & \cdots & \cdots \\ \tau_{1_{t+1}|S_t} & \tau_{2_{t+1}|S_t} & \cdots & \tau_{S_{t+1}|S_t} \end{bmatrix} \qquad \text{Equation 2.8}$$

$$P(X_r = x_r) = \sum_{S_1=1}^{S} \sum_{S_2=1}^{S} \cdots \sum_{S_T=1}^{S} \nu_{S_1} \tau_{S_2|S_1} \tau_{S_3|S_2} \cdots \tau_{S_T|S_{T-1}} \prod_{t=1}^{T} \prod_{i=1}^{I} \pi_{iS_t}^{x_{rit}} \left(1 - \pi_{iS_t}\right)^{1 - x_{rit}} \qquad \text{Equation 2.9}$$

Applications of LTA-based longitudinal DCMs. Recently, a few LTA-based longitudinal DCMs have been evaluated in simulation studies as well as applied in empirical studies. For example, Li et al. (2016) used the LTA with DINA as the measurement model to evaluate the effectiveness of an intervention for four cognitive skills across four measurement occasions for a sample of 109 seventh-grade students. This study provided base-rates of cognitive skills at each measurement occasion and three conditional transition probabilities from Occasion 1 to Occasion 2, Occasion 2 to Occasion 3, and Occasion 3 to Occasion 4, respectively. The results showed that attributes had different base-rates at the beginning and different conditional transition probabilities over time.

Madison (2016) proposed the transitional diagnostic classification model (TDCM) to measure growth in attribute mastery for pretest and posttest data, where the LCDM was adopted as the measurement model along with the LTA as the structural model. A simulation study showed that the TDCM could provide accurate and reliable classification and transition probabilities over time under the variations in the number of attributes, sample size, Q-matrix, pretest and posttest base-rates, and marginal mastery transition probabilities. Additionally, the TDCM was applied to two empirical studies. In both studies, four mathematic skills were assessed before and after an intervention. The results showed that the base-rates of all attributes were improved after the intervention. However, the improvement differed by attributes and the groups, e.g., the control group or the intervention groups.

Higher-order DCM-based Longitudinal DCMs.

Higher-order DCM. Higher-order DCMs (HDCMs) parameterize the structural model of general DCMs in a certain way to reduce the numbers of structural parameters. HDCMs focus on the parameterization in the structural model. Therefore, various types of DCMs (e.g., DINA, DINO, LCDM) could be adopted as the measurement model. On the other hand, as discussed above, several approaches have been utilized to construct the structural model. Since current HDCM-based longitudinal DCMs and the proposed model are parameterized using the logistic regression models, in the following section, HDCMs constructed via a logistic regression model are introduced. More details on constructing HDCMs using other models can be found in studies of Rupp and Templin (2008b), Hartz (2002), and Templin et al. (2008).

de la Torre and Douglas (2004) proposed an HDCM, where a higher-order latent variable was used to account for the correlations among attributes via a logistic regression model. Suppose there are A dichotomous attributes ($\alpha = \alpha_1, \dots, \alpha_A$) that are predicted by a unidimensional latent trait θ; thus, the probability of observing an attribute vector α_r for individual r given θ_r, and the probability of mastering attribute α_{rA} given θ_r could be expressed as follows:

$$P(\alpha_r|\theta_r) = \prod_{\alpha=1}^{A} P(\alpha_{rA}|\theta_r)$$

$$P(\alpha_{rA}|\theta_r) = \frac{\exp(\lambda_{0A} + \lambda_{1A}\theta_r)}{1 + \exp(\lambda_{0A} + \lambda_{1A}\theta_r)}$$

Equation 2.10

where θ_r is a normally distributed latent variable with a mean of 0 and variance of 1, and λ_{0A} and λ_{1A} are the intercept and factor loadings between θ_r and α_A. In Equation 2.10, only a unidimensional higher-order latent variable is included in the model. However, multiple higher-order latent variables are also possible (de la Torre & Douglas, 2004).

HDCM-based longitudinal DCMs. HDCM-based longitudinal DCMs are composites of two model components. The first component is the HDCM, where an higher-order continuous factor, θ_{rt}, is assumed to predict the mastery statuses of multiple lower-order attributes at time t. The second component is the univariate growth curve models (GCMs; e.g., Hoffman, 2015; Raghavarao & Padgett, 2014), which describe the inter- and intra-individual differences in changes of this higher-order factor over T time points.

The HDCM-based longitudinal DCMs are developed in the multilevel model framework, such that the general HDCM-based longitudinal DCMs are three-level models. Suppose A dichotomous attributes ($\alpha = \alpha_1, \ldots, \alpha_A$) have been tested at $t = 1, \ldots, T$ times. Thus, the attribute profile for individual r at time t is represented by $\alpha_{rt} = (\alpha_{r1t}, \ldots, \alpha_{rAt})$.

In Level 1, a DCM evaluates the mastery status of each attribute at each time point. Then, a higher-order DCM is constructed, where a higher-order factor, θ_{rt}, is used to predict the mastery status of lower-order attributes at time t.

In Level 2 and 3, a univariate GCM was adopted to describe the growth of θ_{rt}. In Level 2, θ_{rt} is specified as a function of a random intercept β_{0r} and a random slope β_{1r}, representing the average initial level and growth rate of θ. In Level 3, the random intercept and slope can be predicted by the fixed intercept and slope.

Level 1

$$\pi_{ict} = P(x_{ict}|\alpha_{rt}) \quad \text{Equation 2.11}$$

$$P(\alpha_{rat} = 1|\theta_{rt}) = \frac{\exp[\lambda_{1at}(\theta_{rt} - \lambda_{0at})]}{1 + \exp[\lambda_{1at}(\theta_{rt} - \lambda_{0at})]} \quad \text{Equation 2.12}$$

Level 2

$$\theta_{rt} = \beta_{0r} + \beta_{1r} time_{rt} + \epsilon_{rt}, \epsilon_{rt} \sim MVN(0, R)$$

$$R = \begin{bmatrix} \sigma_1^2 & & \dots & \\ 0 & \sigma_2^2 & \dots & \\ 0 & 0 & \ddots & \\ 0 & 0 & \dots & \sigma_T^2 \end{bmatrix} \quad \text{Equation 2.13}$$

Level 3

$$\beta_{0r} = \gamma_{00} + u_{0r} \quad \text{Equation 2.14}$$

$$\beta_{1r} = \gamma_{01} + u_{1r} \quad \text{Equation 2.15}$$

$$\psi(u_{0r}, u_{1r}) \sim MVN(\binom{0}{0}, G)$$

$$G = \begin{bmatrix} \sigma_{u_0}^2 & \\ \sigma_{u_0}\sigma_{u_1} & \sigma_{u_1}^2 \end{bmatrix} \quad \text{Equation 2.16}$$

In Equation 2.11, π_{irt} represents the probability of a correct item response to item i at time t, given the individual r with attribute profile α_r at time t. Different types of DCMs could be applied to estimate π_{irt}.

In Equation 2.13, $time_{rt}$ represents the time scores at time t. Individuals could have the same time scores or their own time scores at each time point. β_{0r} and β_{1r} are random intercept and random slope, and ϵ_{rt} is the Level 2 residual, which follows a multivariate normal distribution with a mean of 0 and covariance matrix R; this is expressed as $\epsilon_{tr} \sim MVN(0, R)$, where R is a $T \times T$ covariance matrix, and T is the total number of measurement occasions (See Equation 2.13). In addition, time-varying predictors could be included in Level 2, and time-invariant predictors could be included in Level 3.

In Level 3, γ_{00} and γ_{10} represent the fixed initial level and the growth rate, indicating the average initial level and growth rate of θ, respectively. $\mathrm{u_{0r}}$ and $\mathrm{u_{0r}}$ are the individual r's deviations from the average initial level and the average growth rate, which is multivariate normally distributed with means of 0 and covariance matric *G*; this is expressed as $[\mathrm{u_{0i}}, \mathrm{u_{1i}}] \sim \mathrm{MVN}(0, \mathrm{G})$, where *G* is a $P \times P$ covariance matrix for which P is the total number of random effects at Level 3 (See Equation 2.16).

Applications of HDCM-based longitudinal DCMs. Recently, Huang (2017) proposed an HDCM-based longitudinal DCM, where a G-DINA model is used to evaluate mastery status of attributes at each time point. Then, the Rasch model was utilized to construct the higher-order DCM at each time point. Last, a univariate GCM was applied to describe the growth of the higher-order factor over time. In addition, a set of time-invariant predictors (e.g., gender, age) were included in Level 3 to predict the random intercept and slope. This HDCM-based longitudinal DCM was evaluated in three simulation studies which varied several factors, including the sample size, the test length, the number of attributes, the item difficulty, and the number of measurement occasions. The results showed that a large sample size (1000 individuals), enough items (30 items), and more measurement occasions (3 measurement occasions) could improve the parameter recovery and classification accuracy. Additionally, this HCDM-based longitudinal DCM was retrofitted to an empirical testing data, which assessed four attributes in a group of 4,177 high school students across three measurement occasions. The results showed that attributes differed in both the initial base-rates and the amount of improvement of the base-rates, for example, the base-rates of the “geometry” attribute were .90, .89, and .92 across three measurement occasions; however, the base-rates of the “number” attribute were .36, .49, and .58 across three measurement occasions. These results indicated different attributes developed different growth rates.

Limitations of current longitudinal DCMs. Even though the current longitudinal DCMs have provided a few approaches to analyze longitudinal data in the DCM framework; these longitudinal DCMs have limitations that could restrict their usage with empirical data. As discussed above, LTA-based longitudinal DCMs could estimate the changes of attributes directly over time. However, this method required balanced data. In other words, the time interval between measurement occasions cannot be accounted for in the model. This might result in inaccurately estimated transition probabilities if examinees have a different time interval between administrations. On the other hand, HDCM-based longitudinal DCMs estimate the growth of the higher-order factor via the univariate GCM framework, which could cooperate both balanced and unbalanced data. However, HDCM-based longitudinal models measure the growth of higher-order factors to indicate the growth of lower-order attributes, indicating multiple attributes should have similar growth patterns. While empirical studies' demonstrated attributes had different growth patterns, some attributes were improved over time, and some attributes had a nearly consistent base-rate over time. For example, Madison (2016) measured the changes in mastery status for four mathematics skills using pre- and posttest data and found the base-rate of one attribute was almost constant, where the base-rates changed from .65 to .70. However, base-rates of another three attributes improved more, ranging from .38 to .58, .38 to .51, and .59 to .73, respectively. Therefore, it is not reasonable to assume all attributes have the same growth patterns such that the growth of the higher-order factor cannot represent the changes in lower-order attributes well.

Therefore, there is a need to improve the current longitudinal DCMs. The motivation for the current study is to improve the current longitudinal DCMs by developing a multivariate longitudinal DCM, which could incorporate both balanced and unbalanced data, and measure the growth of attributes directly without assuming that attributes have similar growth patterns.

Chapter 3 : Research Design and Methods

In Chapter 3, I describe the proposed multivariate longitudinal DCM, and outline the research questions, the design of two simulation studies, analysis plans, outcome variables, and evaluation criteria.

Proposed Multivariate Longitudinal DCM

The proposed multivariate longitudinal DCM is a composite of two components, the LCDM as the measurement model component that evaluates the mastery status of attributes at each measurement occasion, and a generalized multivariate growth curve model (e.g., GCM; Goldstein, 2011; Hoffman, 2015; MacCallum, Kim, Malarkey, & Kiecolt-Glaser, 1997) as the structural model component that describes the changes of attributes over time via a logistic link function.

Model specification. Let $\mathrm{x_i}$ denote the item response of item i. Only the binary item response was considered in the current study; however, polytomous item responses could be incorporated as well. Let $\mathrm{t} = 1,2, \ldots, \mathrm{T}$ denote the number of measurement occasions; $\mathrm{k} = 1,2, \ldots, \mathrm{K}$ denote the number of attributes; and $\alpha_{\mathrm{rt}}^{\mathrm{k}} = \alpha_{\mathrm{rt}}^{1}, \alpha_{\mathrm{rt}, \ldots,}^{2} \alpha_{\mathrm{rt}}^{\mathrm{k}}$ denote the attribute profile at time t. Only the binary attribute is considered in the current study, however, polytomous attributes could be incorporated as well.

A three-level model is considered in the current study; Level 1 is the item level, Level 2 was the within-person level, and Level 3 is the between-person level.

In Level 1, the LCDM estimates the probability of individual r answering item i correct given profile α_{r} at time t, as shown in Equation 3.1.

In Level 2, $\alpha_{\mathrm{rt}}^{\mathrm{k}}$ represents the mastery status of attribute k at time t, $\mathrm{Time}_{\mathrm{rt}}$ is the time variable for individual i at time t. Then, the log-odds of $\mathrm{P}(\alpha_{\mathrm{rt}}^{\mathrm{k}} = 1)$, indicating the probability of mastery attribute k at time t, are predicted by the random intercept $\beta_{\mathrm{r0}}^{\mathrm{k}}$ and random slope $\beta_{\mathrm{r1}}^{\mathrm{k}}$.

In Level 3, the random intercept $\beta_{0\mathrm{r}}^{\mathrm{k}}$ and random slope $\beta_{1\mathrm{r}}^{\mathrm{k}}$ are predicted by average initial level γ_{00}^{k} and average slope γ_{10}^{k}, respectively. $\mathrm{u}_{0\mathrm{r}}^{\mathrm{k}}$ and $\mathrm{u}_{1\mathrm{r}}^{\mathrm{k}}$ represent the individual $\mathrm{r}'\mathrm{s}$ deviations from the average initial level and growth rate for attribute k.

Level 1 $$\pi_{irt} = P(X_{irt} = 1|\alpha_{rt}) = \frac{\exp(\lambda_{i,0} + \lambda_i^T h(\alpha_{rt}, q_i))}{1 + \exp(\lambda_{i,0} + \lambda_i^T h(\alpha_{rt}, q_i))}$$ Equation 3.1

$$logit\left(P\left(\alpha_{rt}^k = 1\right)\right) = \beta_{r0}^k + \beta_{r1}^k Time_{rt} + \epsilon_{rt}^k$$ Equation 3.2

Level 2

$$R = \begin{bmatrix} \frac{\pi^2}{3} & & \cdots & & \\ 0 & \frac{\pi^2}{3} & \cdots & & \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{\pi^2}{3} & \\ 0 & 0 & \cdots & 0 & \frac{\pi^2}{3} \end{bmatrix}$$ Equation 3.3

$$\beta_{0r}^k = \gamma_{00}^k + u_{0r}^k$$ Equation 3.4

$$\beta_{1r}^k = \gamma_{10}^k + u_{1r}^k$$ Equation 3.5

Level 3

$$G = \begin{bmatrix} \sigma_{u_0}^{(1)2} & & & \\ \sigma_{u_0}^{(1)}\sigma_{u_1}^{(1)} & \sigma_{u_1}^{(1)2} & & \\ \vdots & \ddots & & \\ \sigma_{u_0}^{(1)}\sigma_{u_0}^{(k)} & \sigma_{u_1}^{(1)}\sigma_{u_0}^{(k)} & \sigma_{u_0}^{(k)2} & \\ \sigma_{u_0}^{(1)}\sigma_{u_1}^{(k)} & \sigma_{u_1}^{(1)}\sigma_{u_1}^{(k)} & \sigma_{u_0}^{(k)}\sigma_{u_1}^{(k)} & \sigma_{u_1}^{(k)2} \end{bmatrix}$$ Equation 3.6

As shown in Equation 3.2, ϵ_{rt}^k are the Level 2 residuals, which follow a multivariate normal distribution with means of 0 and TK × TK covariance matrix of R, the diagonal elements are $\frac{\pi^2}{3}$, and off-diagonal elements are fixed to 0, indicating there are no covariances among $\boldsymbol{\epsilon_{rt}}$ across constructs. In Level 3 variance $[u_{0r}^k,\ u_{1r}^k] \sim MVN(0, G)$, G is a $KP \times KP$ covariance matrix, and P is the number of Level 3 random effects.

Research questions.

The purpose of the current study is to develop a multivariate longitudinal DCM and evaluate it under several conditions.

This dissertation aims to answer the following research questions :

(1) Does the proposed model provide satisfied classification accuracy under different conditions?

(2) Do the sample size, the growth patterns, and the number of measurement occasions, the G matrix design, and their interactions impact the item parameter recoveries in the measurement model?

(3) Do the sample size, the growth patterns, and the number of measurement occasions, the G matrix design, and their interactions impact the fixed and random effects recoveries in the generalized growth curve model?

Simulation Study Design

To answer three research questions listed above, a simulation study was conducted, which included four the design factors, (1) the sample size; (2) the growth patterns across attributes; (3) the G matrix design; and (4) the number of measurement occasions. Factors including the Q-matrix, the test length, the initial base-rate, and the item parameters were fixed. Simulation conditions are described below.

Design factors.

Sample size. The current study varied the sample size by 100, 200, and 300 to investigate the requirement for the sample size in the proposed model. Previous simulation studies in longitudinal DCMs used to have a large sample size that normally ranged from 500 to 3000 (e.g., Kaya & Leite, 2017; Madison, 2016; Zhan et al., 2017). However, the empirical studies usually had a relatively smaller sample size, normally ranging from 100 to 400 (e.g., Li et al., 2016; Madison, 2016). Therefore, it was useful to investigate the sufficient sample size for the proposed model to detect the growth of attributes over time, which could guide applied researchers to collect adequate participants without a waste of time and money.

Growth patterns across attributes. The proposed multivariate longitudinal DCM improves the current HDCM-based longitudinal DCMs in its potential for estimating the growth of attributes without assuming that attributes have similar growth trajectories. To examine if the proposed model could measure attributes with different growth patterns and attributes with similar growth patterns equally well, two different growth patterns across attributes were considered in the current study: (1) the even growth pattern in which attributes had similar growth patterns over time and (2) the uneven growth pattern in which attributes had different growth patterns over time.

Figure 3.1 describes these two conditions. Under the even growth pattern condition, the base-rates of all three attributes were improved from the first measurement occasion to the last measurement occasion. Under the uneven growth pattern condition, the base-rates of Attributes 2 and 3 were improved across five measurement occasions, but the base-rates of Attribute 1 kept constant over time.

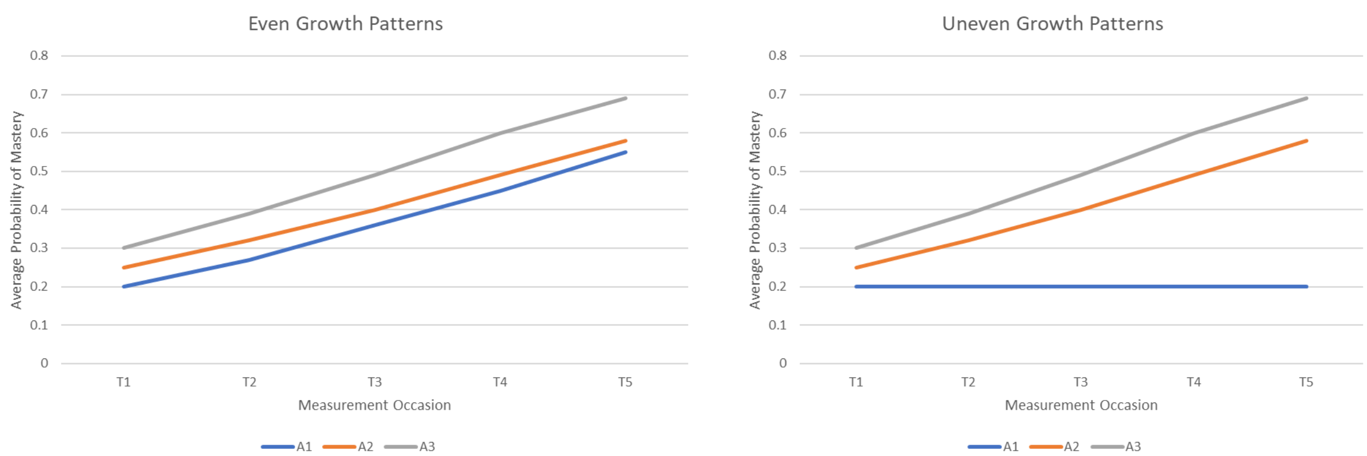

Figure 3.1 *Two patterns of growth across attributes*

Note. T1 to T5 represent the first to the fifth measurement occasion; A1, A2, and A3 represent Attribute 1, Attribute 2, and Attribute 3, respectively.

G matrix design. The G matrix plays an important role in the multivariate GCM, which reflects the relationships between outcomes across time. It is one of the main interests in the longitudinal studies that measure multiple outcomes over time (e.g., Hoffman, 2015).

To examine if the proposed multivariate longitudinal DCM can detect the relationships among attributes, two types of G matrices are considered in the current study: (1) under the equal correlation condition, all attributes had equal correlations between intercept, slopes, and intercept and slope, meaning that attributes are equally correlated, and (2) under the unequal correlation condition, as described in Figure 3.1, Attribute 2 and Attribute 3 had equal correlations between intercept, slopes, and intercept and slope, but

Attribute 1 had lower correlations with Attribute 2 and 3. Table 3.1 presents the two types of G matrices and corresponding correlation matrices.

Table 3.1 G Matrix Specification and Corresponding Correlation Matrix

G Matrix Specification and Corresponding Correlation Matrix

	Equal correlation condition							Unequal correlation condition					
Correlation Matrix													
	u_0^1	u_1^1	u_0^2	u_1^2	u_0^3	u_1^3		u_0^1	u_1^1	u_0^2	u_1^2	u_0^3	u_1^3
u_0^1	**1.0**						u_0^1	**1.0**					
u_1^1	.20	**1.0**					u_1^1	.10	**1.0**				
u_0^2	.90	.10	**1.0**				u_0^2	.90	.01	**1.0**			
u_1^1	.10	.25	.20	**1.0**			u_1^1	.01	.01	.20	**1.0**		
u_0^3	.90	.10	.90	.10	**1.0**		u_0^3	.10	.01	.90	.10	**1.0**	
u_1^3	.10	.25	.10	.25	.20	**1.0**	u_1^3	.01	.01	.10	.25	.20	**1.0**
Covariance Matrix													
	$\sigma^2_{u_0^1}$	$\sigma^2_{u_1^1}$	$\sigma^2_{u_0^2}$	$\sigma^2_{u_1^2}$	$\sigma^2_{u_0^3}$	$\sigma^2_{u_1^3}$		$\sigma^2_{u_0^1}$	$\sigma^2_{u_1^1}$	$\sigma^2_{u_0^2}$	$\sigma^2_{u_1^2}$	$\sigma^2_{u_0^3}$	$\sigma^2_{u_1^3}$
$\sigma^2_{u_0^1}$	**.1500**						$\sigma^2_{u_0^1}$	**.1500**					
$\sigma^2_{u_0^1}$	.0173	**.0500**					$\sigma^2_{u_0^1}$	.0173	**.0500**				
$\sigma^2_{u_0^2}$	.1350	.0087	**.1500**				$\sigma^2_{u_0^2}$	.1350	.0009	**.1500**			
$\sigma^2_{u_0^2}$	.0087	.0125	.0173	**.0500**			$\sigma^2_{u_0^2}$	.0009	.0005	.0173	**.0500**		
$\sigma^2_{u_0^3}$	.1350	.0087	.1350	.0087	**.1500**		$\sigma^2_{u_0^3}$	.1350	.0009	.1350	.0087	**.1500**	
$\sigma^2_{u_0^3}$	.0087	.0125	.0087	.0125	.0173	**.0500**	$\sigma^2_{u_0^3}$	.0009	.0005	.0087	.0125	.0173	**.0500**

Note. u_0^k and u_1^k represent the random intercept and slope for attributes; $\sigma^2_{u_0^k}$ and $\sigma^2_{u_0^k}$ represent the random intercept and slope variance for attributes.

Number of measurement occasions. Previous simulation studies in HDCM-based longitudinal DCMs showed inconsistent results in the impacts of the number of measurement occasions on the classification accuracy. Huang (2017) found the number of measurement occasions (e.g., 2 or 3 measurement occasions) did not influence the classification accuracy significantly. However, Zhan et al. (2017) found the classification accuracy slightly increased as the number of measurement occasions increased. For the growth model, more measurement occasions are associated with good parameter recoveries (e.g., Preacher, Wichman, MacCallum, & Briggs, 2008). To examine whether the number of measurement occasions impacted the performance of the proposed multivariate longitudinal DCM, the number of measurement occasions varied between 3 and 5 in the current study.

Fixed conditions.

Test length. A test of 30 binary items was simulated in the current study. The test length fell within the range of applied research as well as simulation studies in the longitudinal DCMs (e.g., Huang, 2017; Kaya & Leite, 2017; Madison, 2016)

Q-matrix. As discussed above, DCMs are able to incorporate both the simple structure and the complex structure of the Q-matrix. In the current study, a complex structure of the Q-matrix was specified as shown in Table 3.2. Each item measures up to two attributes and attributes were assessed by equal numbers of items. This Q-matrix design was suggested by previous applied research and simulation studies (e.g., Bradshaw et al., 2014; Bradshaw & Templin, 2014; Kaya & Leite, 2017; Madison, 2016).

Table 3.2

Q-matrix Design

Item	Attribute 1	Attribute 2	Attribute 3	Item	Attribute 1	Attribute 2	Attribute 3
1	1	0	0	16	1	1	0
2	0	1	0	17	1	0	1
3	0	0	1	18	0	1	1
4	1	1	0	19	1	0	0
5	1	0	1	20	0	1	0
6	0	1	1	21	0	0	1
7	1	0	0	22	1	1	0
8	0	1	0	23	1	0	1
9	0	0	1	24	0	1	1
10	1	1	0	25	1	0	0
11	1	0	1	26	0	1	0
12	0	1	1	27	0	0	1
13	1	0	0	28	1	1	0
14	0	1	0	29	1	0	1
15	0	0	1	30	0	1	1

Initial Base-rates. The initial base-rate was fixed to .20, .25, and .30 for Attribute 1, Attribute 2, and Attribute 3, respectively. The previous empirical studies on measuring growth of attributes found initial base-rates ranged from .02 to .90 and suggested an easier attribute might have a base-rate approximately .60, a medium attribute might have a base-rate approximately 0.40, and a hard attribute might have a base-rate approximately .20 (Madison, 2016); therefore, the base-rates are set to .20, .25, and .30 to mimic the hard, medium-hard, and medium attributes at the first measurement occasion.

Fixed effects ($\gamma_{00}^{k}, \gamma_{01}^{k}$). The linear growth of the log-odds of the probability of mastering attributes was considered in the current study. It should be noted that the linear growth of the log-odds of the probability did not necessarily result in the linear growth of base-rates over time. Table 3.3 presents the fixed effects under both even and uneven growth pattern conditions.

Table 3.3

Initial Level and Growth Rates of Linear Predictors

	Even Growth Patterns			Uneven Growth Patterns		
	A1	A2	A3	A1	A2	A3
γ_{00}	-1.38	-1.10	-.85	-1.38	-1.10	-.85
γ_{01}	.05	.04	.05	0	.04	.05

Note. A1, A2, and A3 represent Attribute 1, Attribute 2, and Attribute 3.

Time variables. The current study planned to mimic the context of the interim assessments, which are administered several times within a school year (Great Schools Partnership, 2013). The common interval ranges from 6-8 weeks, such that individuals might receive the assessment at different times. Therefore, the current study set the time interval to 8 weeks and the unit of time to 1 week. The mean and standard deviation of time variables at each measurement occasion was fixed to $\mu_{time} = (0, 8, 16, 24, 32)$ and $\sigma_{time} = 1$, such that each individual had his/her own time variable at each measurement occasion to mimic the unbalanced data design.

Table 3.4 presents the average base-rates of attributes across five measurement occasions, which was obtained by using the mean of the time variable and fixed effects shown in Table 3.3. Under the even growth pattern condition, the probabilities of mastery of three attributes were improved by .35, .38, and .39, respectively, across the time, and under the uneven pattern condition, the base-rate of Attribute 1 had a constant of .20, and the probabilities of mastery were improved by .38 and .39 for Attributes 2 and 3, respectively. This amount of improvement fell in the range of improvement of base-rates found in the previous studies (Li et al., 2016; Madison, 2016).

Table 3.4

Base-rates of Attributes Over Time

	T1	T2	T3	T4	T5
Even growth pattern					
A1	.20	.27	.36	.45	.55
A2	.25	.32	.40	.49	.58
A3	.30	.39	.49	.60	.59
Uneven growth pattern					
A1	.20	.20	.20	.20	.20
A2	.25	.32	.40	.49	.58
A3	.30	.39	.49	.60	.59

Note. A1, A2, and A3 represent Attribute 1, Attribute 2, and Attribute 3; T1 to T5 represent the first measurement occasion to the fifth measurement occasion.

Item parameters. The intercepts of all items were fixed into −1.5 indicating the probability of having a correct answer was .18. The simple main effects of all items were fixed into 1.5, indicating the probability of having a correct answer was .50 given mastering this attribute. The interaction effects between two attributes were fixed to .50, indicating the probability of having a correct answer was .88, given mastering two attributes.

Data generation procedures.

Data were generated in R, version 3.4.2 (R Core Team, 2017). Each condition was replicated 100 times.

Data generation procedures included two stages: first, I generated the probability of mastery for each attribute at five measurement occasions, then generate the mastery statuses of them; lastly, I generated item response data, which are proceeded as follows:

(1) Generate the linear predictors of the probability of mastery for each attribute by using the intercept and slope parameters, time variables, and G matrix for each individual;

(2) Convert this linear predictor into the probability of mastery;

(3) A binary mastery status for each attribute is randomly drawn from the binomial distribution with the probability of mastering attributes.

(4) Generate the probability of having a correct answer for each item using a prespecified Q-matrix, item parameters, and person profiles.

(5) A binary item response is randomly sampled from the binomial distribution with the probability obtained from the last step.

Analysis Plan and Outcome Variables

A Markov Chain Monte Carlo (MCMC) algorithm was adopted to estimate model parameters, which was implemented in the JAGS software (Plummer, 2003) by using the R2jags package (Su & Yajima, 2015) in the programming environment R (R Development Core Team, 2017). The JAGS syntax and more details of MCMC analyses can be found in Appendix.

The LCDM was applied to estimate the mastery statuses of attributes at each measurement occasion. For example, as described in the Q-matrix in Table 3.2, item 4 measured both Attribute 1 and Attribute 2. Thus, the probability of providing a correct answer to item 4 given the latent class c at Time t can be expressed as follows:

$$\pi_{4ct} = P(x_{4ct} = 1|\alpha_{ct}) = \frac{\exp(\lambda_{4,0} + \lambda_{4,1,(1)}(\alpha_1) + \lambda_{4,1,(2)}(\alpha_2) + \lambda_{4,2,(1,2)}(\alpha_1 \cdot \alpha_2))}{1 + \exp(\lambda_{4,0} + \lambda_{4,1,(1)}(\alpha_1) + \lambda_{4,1,(2)}(\alpha_2) + \lambda_{4,2,(1,2)}(\alpha_1 \cdot \alpha_2))} \quad \text{Equation 3.7}$$

For items that only measure one attribute, only the intercept and the main effect of this item were included in the equation.

The generalized multivariate GCM was applied to measure the changes in mastery statuses of attributes over time. First, as suggested by MacCallum et al., (1997), Curran, McGinley, Serrano, and Burfeind (2012) and Hoffman (2015), a synthesized variable was created, which was a composite of multiple outcome variables (α_{rt}^{k} in the current study), then a series of dummy variables as exogenous predictors were adopted to control which specific outcomes were referenced within different parts of the model. Let dv_{rt} denote the synthesized variable, which contained individual $r's$ mastery statuses for three attributes across four measurement occasions. A total of three dummy variables, $A1, A2,$ and $A3,$ were included in the model to distinguish which specific element belonged to which specific outcome variables, where $A1$ was equal to 1 for Attribute 1 and $A1$ was equal to 0 for other attributes. Therefore, the probability of mastering attribute α_{rt}^{k} $(k = 1,2,3)$ at time t could be described as follows:

$$
\begin{aligned}
&logit\big(P(dv_{rt} = 1)\big) \\
&= A1[(\gamma_{00}^{1} + u_{0r}^{1}) + (\gamma_{10}^{1} + u_{1r}^{1})Time_{rt}] \\
&\quad + A2[(\gamma_{00}^{2} + u_{0r}^{2}) + (\gamma_{10}^{2} + u_{1r}^{2})Time_{rt}] \\
&\quad + A3[(\gamma_{00}^{3} + u_{0r}^{3}) + (\gamma_{10}^{3} + u_{1r}^{3})Time_{rt}]
\end{aligned}
\qquad \text{Equation 3.8}
$$

where the main effects of $A1, A2$ and $A3$ represent the initial levels for three attributes, and the interaction effects between dummy variables and time scores represent the growth rates for attributes.

Once data analysis was finished, the following outcome variables across all 100 replications were obtained for all conditions:

(1) Gelman-Rubin diagnostic $(\hat{R})$ of parameters, including item parameters in the LCDM and both fixed effects and random effects parameters in the generalized growth curve model.

(2) The distribution of estimated parameters, including the mean, standard deviation, and quantiles.

Evaluation Criteria

Convergence rates, the classification accuracy of attributes at each measurement occasion, and the parameter recovery were evaluated in the current study to examine the performance of the proposed model under different conditions.

Convergence rates.

Cnvergence was assessed by using the Gelman-Rubin diagnostic $(\hat{R})$, also referred to as the “potential scale reduction factor” (Gelman & Rubin, 1992). Suppose there are m independent Markov chains, $\hat{R}$ is given by:

$$\sqrt{\hat{R}} = \sqrt{\frac{n-1}{n} + \frac{1}{n}\frac{B}{W}} \qquad \text{Equation 3.9}$$

where B is the variance between the means of the m chains, W is the average of the m within-chain variances, and n is the number of iterations of the chain after discarding the iterations as burn-in. If the algorithm converges, $\hat{R}$ is approaching 1, indicating a stationary distirbution has been achieved because the marginal posterior variance (weighted combo of between and within- chain variance) are equal to the within-chain variances. In the current study, $\hat{R}$ was calculated for all model parameters, and I adopted the criteria of $\hat{R} < 1.2$ as the indicator of convergence as suggested by the previous study (e.g., Sinharay, 2003).

In one replication, if one or more parameters had the $\hat{R}$ larger than 1.2, this replication was regarded as non-converged. After a total of 100 replications, the convergence rates for this condition was calcualted and reported. Only the results from the converged replications were kept and used in the following analysis.

Classification accuracy.

The classification accuracy was evaluated by using (1) the bias of estimated probability of attribute mastery, (2) the correct classification rates for each mastery status, and (3) Cohen's kappa (Cohen, 1960).

Bias of estimated probability of attribute mastery was the difference between the estimated and the true probability of attribute matery. The correct classifcation rates for each mastery status included (1) the corrected classification rates for individuals who truly mastered one attribute, and (2) the corrected classification rates for individuals who truly did not master one attribute. Cohen's kappa measures the aggreement between the true and the estimated mastery status.

The estimated class membership was obtained by applying 0.5 as the cutpoint, meaning that an individual with an estimated probability larger than 0.5 would be classified as mastery, vice versa.

Parameter recovery.

The bias and mean squared error (MSE) of estimated parameters, including item parameters from the measurement model, intercept and slope parameters, and variance and covariance parameters from the structural model were computed to assess the parameter recovery in each condition.

$$Bias_{\theta} = \frac{\sum_{r=1}^{R}\sum_{i}^{N}(\hat{\theta}_{ir} - \theta_i)}{RN} = \bar{\hat{\theta}}_{ir} - \theta_i \qquad \text{Equation 3.10}$$

$$MSE_{\theta} = \frac{\sum_{r=1}^{R}\sum_{i=1}^{N}\left(\hat{\theta}_{ir} - \theta_i\right)^2}{RN} \qquad \text{Equation 3.11}$$

where θ represents the estimated parameter, which is the mean of the sample distribution obtained from the Baysian estimation. R is the number of replications; N is the number of elements in the set of θ.

A factorial analysis of variance was adopted to assess the impact of design factors on outcome variables. In all analyses, the α level was controlled at .05 level, and partial η^2 was adopted as the measure of effect sizes. According to Cohen's (1988) convention, partial η^2 values of .01, .06, and .14 were regarded as small, medium, and large effects.

Chapter 4 : Results

Convergence Rates

As indicated in the previous chapter, the Gelman-Rubin diagnostic ($\hat{R}$) of item parameters in the LCDM, fixed effects and randon effects parameters in the generalized growth curve model were evaluated, and I adopted the criteria of $\hat{R} < 1.2$ as the indicator of convergence as suggested by the previous study (e.g., Sinharay, 2003). When all the parameters that included the item parameters in the LCDM, fixed effects and the random effects parameters in the generalized growth curve model were converged in one replication, this replication was regarded as converged. Table 4.1 presents the average convergence rates under different conditions, which shows the average convergence rate is 0.95 under the conditions with three measurement occasions ($MO = 3$). And, the average convergence rate is 0.97 under the conditions with five measurement occasions ($MO = 5$). Only the converged replications were used in the following analyses.

Table 4.1

Average Convergence Rates

		Three Measurement Occasions ($MO = 3$)			Five Measurement Occasions ($MO = 5$)		
		N100	N200	N300	N100	N200	N300
G1	gam1	.91	.97	.96	.98	1	.90
	gam2	.88	.95	.91	.98	.99	.95
G2	gam1	.93	.97	.99	.98	.98	.92
	gam2	.91	.97	.99	1	1	1

Note. N100, N200, and N300 represent the sample size of 100, 200, and 300, respectively; G1 and G2 represent equal correlation G and unequal correlation conditions of G matrix, respectively; gam1 and gam2 represent the same growth pattern across attributes and unequal growth patterns across attributes respectively.

Classification Accuracy

The classification accuracy was evaluated by using (1) bias of the estimated probability of attribute mastery, (2) correct classification rates for each mastery status, and (3) Cohen's kappa.

Table 4.2 presents the average bias of probability of attribute mastery under the conditions when $MO = 3$, which shows the probability of attribute mastery was recovered well under most conditions.

Similar patterns were found when $MO = 5$. The average bias of the probability of attribute mastery was all closed to 0 under most conditions as shown in Table 4.3.

Table 4.2

Bias of Probability of Attribute Mastery (MO=3)

		T1			T2			T3		
		A1	A2	A3	A1	A2	A3	A1	A2	A3
gam1	N100	.	.	.	-.01	.	-.01	.	.	.
	N200	.	.	.	-.01	-.01	.	.	.	.01
	N300	.	.	.	.	-.01	.	.	.	-.01
gam2	N100	.	.	.	-.01	-.01	-.01	-.01	.	-.01
	N200	-.01	.	.	-.01	.	.	-.01	.	.
	N300	.	.	.	-.01	-.01	.	.	-.01	.
gam1	N100	.	.	-.01	-.01	-.01	-.01	-.01	.	.
	N200	.	.	.	-.01	.	.	.	.	.
	N300	.	.	.	.	.	.	.	.	.
gam2	N100	-.01	.	-.01	-.01	.	-.01	-.01	.	.
	N200	.	.	-.01	-.01	.	-.01	.	.	.
	N300	.	.	.01	-.01	-.01	.	.	.	.

Note. T1 to T3 represent the first to the third measurement occasion; A1, A2, and A3 represent Attribute 1, Attribute 2, and Attribute 3; N100, N200, and N300 represent the sample size of 100, 200, and 300, respectively; G1 and G2 represent equal correlation G and unequal correlation conditions of G matrix, respectively; gam1 and gam2 represent the same growth pattern across attributes and unequal growth patterns across attributes, respectively; . represents < .001.

Table 4.3

Bias of Probability of Attribute Mastery (MO=5)

			T1			T2			T3			T4			T5		
			A1	A2	A3	A1	A2	A3	A1	A2	A3	A1	A2	A3	A1	A2	A3
G1	gam1	N100	.	.	.01	.	.	.	.	.	.	.	.	.	.	.	.
		N200	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
		N300	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	gam2	N100	.	.	.01	.	.	.	.	-.01	.	.	.	.	.	.	.
		N200	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
		N300	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
G2	gam1	N100	.	.01	.	.	.	.	.	.	.	.	.	.	.	.	.
		N200	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
		N300	.	.01	.	.	.	.	.	.	.	.	.	.	.	.	.
	gam2	N100	.01	.	.01	.	-.01	.	.	-.01	.	.	-.01	.	.	.	.
		N200	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
		N300	.	.	.	.	.	.	.	.	.	.	.	.	.01	.02	.01

Note. T1 to T5 represent the first to the fifth measurement occasion; A1, A2, and A3 represent Attribute 1, Attribute 2, and Attribute 3; N100, N200, and N300 represent the sample size of 100, 200, and 300, respectively; G1 and G2 represent equal correlation G and unequal correlation conditions of G matrix, respectively; gam1 and gam2 represent the same growth pattern across attributes and unequal growth patterns across attributes, respectively; . represents <.001.

Table 4.4 presents the average correct classification rates for individuals who truly mastered attributes, and Table 4.5 presents the correct classification rates for individuals who truly did not master attributes under different conditions when $MO = 3$. The average correct classication rates were very low for individuals who truly mastered the attributes at the first measurement occasion ($T = 1$), but the correct classification rates improved as the number of measurement occasions increased as shown in Table 4.4. For individuals who truly did not master the attributes, Table 4.5 shows that the correct classification rates were perfect at the first measurement occasion, and then decreased to about 0.9 at the following two measurement occasions.

This pattern might be due to the estimated probability of attribute mastery were very low at the first measurement occasion; the majority of individuals' probabilities were lower than 0.5. After I applied 0.5 as the cutpoint to classify individuals into mastery or non-mastery classes, most of the individuals were classified into the non-mastery class, even they truly mastered the attributes.

As shown in Table 4.6 and Table 4.7, similar patterns were found when $MO = 5$. In summary, even though the probability of attribute mastery were recovered well, the correct classification rates depended on the individuals' mastery status and the cutpoint that was adopted to classify individuals.

Table 4.4

Average Correct Classification Rates for Individuals Who Truly Mastered Attribute (MO=3)

			T1			T2			T3		
			A1	A2	A3	A1	A2	A3	A1	A2	A3
G1	gam1	N100	.04	.06	.08	.61	.65	.70	.86	.88	.89
		N200	.03	.03	.04	.63	.69	.75	.87	.89	.91
		N300	.02	.02	.03	.64	.69	.73	.88	.88	.89
	gam2	N100	.06	.05	.08	.53	.66	.71	.80	.87	.88
		N200	.03	.03	.04	.52	.69	.73	.79	.88	.90
		N300	.04	.02	.02	.55	.68	.74	.81	.88	.91
G2	gam1	N100	0	.06	.11	.62	.64	.73	.85	.87	.89
		N200	.02	.02	.04	.64	.67	.73	.87	.88	.90
		N300	.02	.02	.03	.66	.70	.74	.88	.89	.91
	gam2	N100	.04	.06	.06	.50	.66	.70	.78	.86	.87
		N200	.02	.02	.03	.54	.69	.71	.81	.88	.90
		N300	.02	.02	.02	.55	.69	.74	.82	.89	.90

Note. T1 to T3 represent the first to the third measurement occasion; A1, A2, and A3 represent Attribute 1, Attribute 2, and Attribute 3; N100, N200, and N300 represent the sample size of 100, 200, and 300, respectively; G1 and G2 represent equal correlation G and unequal correlation conditions of G matrix, respectively; gam1 and gam2 represent the same growth pattern across attributes and unequal growth patterns across attributes, respectively.

Table 4.5

Average Correct Classification Rates for Individuals Who Truly did not Mastered Attribute (MO=3)

			T1			T2			T3		
			A1	A2	A3	A1	A2	A3	A1	A2	A3
G1	gam1	N100	1	1	.99	.94	.91	.85	.93	.92	.91
		N200	1	1	1	.94	.90	.84	.94	.93	.91
		N300	1	1	1	.94	.90	.85	.94	.93	.92
	gam2	N100	1	1	1	.96	.90	.84	.96	.92	.90
		N200	1	1	1	.97	.90	.84	.97	.92	.92
		N300	1	1	1	.97	.90	.84	.97	.94	.91
G2	gam1	N100	1	1	.99	.95	.90	.86	.94	.93	.90
		N200	1	1	1	.95	.91	.84	.94	.93	.91
		N300	1	1	1	.95	.90	.84	.93	.93	.92
	gam2	N100	1	1	1	.97	.90	.86	.96	.93	.91
		N200	1	1	1	.97	.90	.85	.96	.92	.91
		N300	1	1	1	.97	.91	.84	.96	.93	.92

Note. T1 to T3 represent the first to the third measurement occasion; A1, A2, and A3 represent Attribute 1, Attribute 2, and Attribute 3; N100, N200, and N300 represent the sample size of 100, 200, and 300, respectively; G1 and G2 represent equal correlation G and unequal correlation conditions of G matrix, respectively; gam1 and gam2 represent the same growth pattern across attributes and unequal growth patterns across attributes, respectively.

Table 4.6

Average Correct Classification Rates for Individuals Who Truly Mastered Attribute (MO=5)

			T1			T2			T3			T4			T5		
			A1	A2	A3	A1	A2	A3	A1	A2	A3	A1	A2	A3	A1	A2	A3
G1	gam1	N100	0	.04	.06	.69	.73	.76	.86	.88	.88	.90	.91	.92	.92	.94	.94
		N200	.03	.02	.03	.70	.73	.77	.86	.87	.89	.91	.91	.92	.93	.93	.94
		N300	.02	.02	.03	.70	.73	.77	.87	.87	.88	.91	.91	.91	.93	.93	.94
	gam2	N100	.04	.04	.07	.60	.71	.76	.80	.86	.89	.87	.90	.92	.90	.92	.93
		N200	0	.02	.03	.62	.73	.77	.82	.86	.89	.88	.91	.92	.90	.93	.94
		N300	.01	.02	.02	.64	.73	.77	.82	.88	.88	.88	.91	.92	.90	.93	.94
G2	gam1	N100	.05	.08	.07	.69	.73	.75	.87	.87	.88	.91	.91	.91	.93	.94	.93
		N200	0	.03	.04	.70	.73	.76	.86	.87	.88	.90	.91	.92	.92	.93	.94
		N300	.02	.02	.03	.69	.73	.76	.86	.86	.88	.90	.90	.91	.91	.92	.93
	gam2	N100	.05	.04	.07	.64	.70	.75	.82	.86	.87	.88	.91	.92	.90	.92	.94
		N200	.02	.02	.04	.63	.73	.77	.83	.87	.89	.87	.91	.91	.90	.93	.94
		N300	.02	.02	.02	.64	.74	.76	.83	.87	.88	.88	.91	.91	.90	.93	.93

Note. T1 to T3 represent the first to the third measurement occasion; A1, A2, and A3 represent Attribute 1, Attribute 2, and Attribute 3; N100, N200, and N300 represent the sample size of 100, 200, and 300, respectively; G1 and G2 represent equal correlation G and unequal correlation conditions of G matrix, respectively; gam1 and gam2 represent the same growth pattern across attributes and unequal growth patterns across attributes, respectively.

Table 4.7

Average Correct Classification Rates for Individuals Who Truly did not Mastered Attribute (MO=5)

			T1			T2			T3			T4			T5		
			A1	A2	A3	A1	A2	A3	A1	A2	A3	A1	A2	A3	A1	A2	A3
G1	gam1	N100	1	1	1	.87	.83	.79	.88	.87	.85	.92	.90	.89	.94	.94	.93
		N200	1	1	1	.86	.83	.79	.87	.87	.86	.92	.91	.90	.94	.94	.93
		N300	1	1	1	.86	.82	.79	.87	.87	.86	.91	.91	.90	.94	.94	.93
	gam2	N100	1	1	.99	.91	.84	.79	.91	.87	.86	.94	.91	.90	.96	.94	.93
		N200	1	1	1	.91	.83	.79	.91	.87	.85	.94	.91	.89	.95	.93	.93
		N300	1	1	1	.90	.82	.78	.90	.87	.85	.93	.91	.90	.96	.94	.93
G2	gam1	N100	1	1	.99	.86	.82	.80	.86	.86	.85	.91	.90	.90	.94	.93	.93
		N200	1	1	1	.85	.82	.79	.87	.86	.85	.91	.91	.89	.94	.94	.93
		N300	1	1	1	.85	.82	.79	.86	.85	.85	.91	.90	.89	.93	.92	.92
	gam2	N100	1	1	.99	.90	.85	.80	.90	.86	.85	.94	.91	.90	.96	.94	.94
		N200	1	1	1	.90	.82	.79	.90	.87	.86	.94	.92	.90	.95	.94	.93
		N300	1	1	1	.89	.82	.79	.90	.87	.85	.93	.91	.89	.94	.92	.91

Note. T1 to T3 represent the first to the third measurement occasion; A1, A2, and A3 represent Attribute 1, Attribute 2, and Attribute 3; N100, N200, and N300 represent the sample size of 100, 200, and 300, respectively; G1 and G2 represent equal correlation G and unequal correlation conditions of G matrix, respectively; gam1 and gam2 represent the same growth pattern across attributes and unequal growth patterns across attributes, respectively.

Cohen's kappa was calculated to evaluate the degree of agreement between the estimated and true mastery status. Table 4.8 presents the average kappa under different conditions when $MO = 3$. The calucation of kappa required that both true and estimated mastery status should had at least two levels; however, estimated mastery status only had one level under some conditions, espeically at the first measurement occasion. Therefore, kappa were not applicable under some conditions. Results found that kappa values improved as the time increased. This pattern might be due to the same reason as discussed above that the estimated probability of mastery was very low for all individuals at the first and second meaurement occasions, such that after applying 0.5 as the cutpoint, the most of individuals who truly mastered the attributes were falsely classified to non-mastery. Therefore, kappa values were low at the beginning but improved as the number of measurement occasions increased. Similar patterns were found when $MO = 5$.

In summary, the agreement between true and estimated mastery status improved as the number of measurement occasions increased, and it was influenced by the cutpoint applied to classify individuals.

Table 4.8

Average Kappa (MO=3)

		T1			T2			T3		
		A1	A2	A3	A1	A2	A3	A1	A2	A3
gam1	N100	.	.	.	.58	.58	.55	.80	.80	.80
	N200	.	.	.	.60	.60	.59	.81	.83	.82
	N300	.	.	.	.61	.60	.58	.83	.82	.82
gam2	N100	.	.	.	.55	.57	.56	.79	.79	.79
	N200	.	.	.	.56	.60	.57	.79	.81	.82
	N300	.	.	.	.58	.60	.58	.80	.82	.82
gam1	N100	.	.	.	.	.57	.59	.80	.80	.80
	N200	.	.	.	.63	.60	.58	.82	.81	.81
	N300	.	.	.	.64	.62	.58	.82	.82	.83
gam2	N100	.	.	.	.53	.58	.56	.76	.80	.79
	N200	.	.	.	.58	.61	.57	.79	.80	.80
	N300	.	.	.	.58	.62	.58	.81	.82	.82

Note. T1 to T3 represent the first to the third measurement occasion; A1, A2, and A3 represent Attribute 1, Attribute 2, and Attribute 3; N100, N200, and N300 represent the sample size of 100, 200, and 300, respectively; G1 and G2 represent equal correlation G and unequal correlation conditions of G matrix, respectively; gam1 and gam2 represent the same growth pattern across attributes and unequal growth patterns across attributes, respectively; . presents the kappa for this condition was not applicable.

Table 4.9

Average Kappa (MO=5)

			T1			T2			T3			T4			T5		
			A1	A2	A3	A1	A2	A3	A1	A2	A3	A1	A2	A3	A1	A2	A3
G1	gam1	N100	.	.	.	.57	.55	.55	.74	.75	.72	.82	.81	.81	.87	.88	.86
		N200	.	.	.	.56	.56	.56	.73	.74	.74	.82	.82	.82	.87	.87	.87
		N300	.	.	.	.56	.55	.56	.73	.73	.74	.83	.82	.82	.87	.87	.87
	gam2	N100	.	.	.	.54	.55	.55	.72	.72	.75	.82	.81	.82	.86	.86	.86
		N200	.	.	.	.56	.56	.56	.73	.73	.73	.82	.82	.81	.86	.86	.87
		N300	.	.	.	.55	.55	.55	.73	.74	.74	.82	.82	.82	.86	.87	.87
G2	gam1	N100	.	.	.	.56	.55	.55	.72	.72	.73	.82	.81	.81	.87	.86	.87
		N200	.	.	.	.56	.55	.55	.73	.73	.74	.82	.82	.81	.87	.87	.86
		N300	.	.	.	.54	.55	.55	.72	.71	.73	.80	.80	.80	.85	.84	.85
	gam2	N100	.	.	.	.56	.55	.55	.72	.72	.73	.82	.82	.82	.86	.86	.87
		N200	.	.	.	.54	.55	.55	.73	.74	.75	.81	.83	.82	.85	.87	.87
		N300	.	.	.	.54	.55	.54	.73	.74	.73	.81	.81	.80	.	.	.

Note. T1 to T5 represent the first to the fifth measurement occasion; A1, A2, and A3 represent Attribute 1, Attribute 2, and Attribute 3; N100, N200, and N300 represent the sample size of 100, 200, and 300, respectively; G1 and G2 represent equal correlation G and unequal correlation conditions of G matrix, respectively; gam1 and gam2 represent the same growth pattern across attributes and unequal growth patterns across attributes, respectively; ; . presents the kappa for this condition was not applicable.

Parameter Recovery

The bias and mean square error (MSE) of the estimated parameters were computed to assess the parameter recovery in each condition through the simulation. Then, ANOVA tests were conducted to assess the impact of the design factors on the bias and MSE values of the estimated parameters of the measurement model and the structural model, respectively.

Measurement model parameter recovery.

There were three sets of item parameters in the LCDM: the intercept (λ_0), the main effect (λ_{α_k}), and the interaction effect ($\lambda_{\alpha_k \alpha_{k'}}$) parameters. Therefore, the average bias and MSE of all three sets of item parameters were assessed to evaluate the measurement model parameter recoveries.

As presented in Table 4.10, the proposed the model achieved good parameter recoveries in intercept and main effect parameters, but the interaction parameters had relatively large bias and MSE values under most conditions. However, the recovery of interaction effect parameters was improved as the sample size and the number of measurement occasions increased.

Table 4.10

Summary of Measurement Model Parameter Recoveries

			Three Measurement Occasions (MO=3)						Five Measurement Occasions (MO=5)					
			λ_0		λ_{α_k}		$\lambda_{\alpha_k \alpha_{k'}}$		λ_0		λ_{α_k}		$\lambda_{\alpha_k \alpha_{k'}}$	
			Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
G1	gam1	N100	.03	.05	-.09	.13	.58	.60	.03	.03	-.07	.07	.31	.24
		N200	.02	.03	-.06	.07	.32	.27	.00	.02	-.03	.04	.15	.11
		N300	.02	.02	-.04	.04	.23	.17	.01	.01	-.03	.03	.11	.07
	gam2	N100	.04	.05	-.08	.13	.62	.68	.02	.03	-.07	.07	.33	.27
		N200	.03	.03	-.06	.07	.36	.31	.01	.02	-.03	.04	.16	.11
		N300	.01	.02	-.04	.04	.24	.19	.00	.01	-.02	.03	.09	.07
G2	gam1	N100	.04	.05	-.09	.13	.62	.66	.03	.04	-.07	.07	.31	.25
		N200	.02	.03	-.05	.07	.36	.32	.02	.02	-.04	.04	.17	.11
		N300	.01	.02	-.03	.05	.25	.20	.01	.01	-.02	.03	.10	.07
	gam2	N100	.03	.05	-.08	.13	.66	.73	.02	.03	-.06	.07	.34	.28
		N200	.03	.03	-.06	.07	.39	.36	.02	.02	-.04	.04	.18	.13
		N300	.01	.02	-.03	.05	.28	.26	.01	.01	-.02	.03	.11	.08

Note. λ_0, λ_{α_k}, and $\lambda_{\alpha_k \alpha_{k'}}$ represents the intercept, main effect, and interaction effect parameters of the LCDM; A1, A2, and A3 represent Attribute 1, Attribute 2, and Attribute 3; N100, N200, and N300 represent the sample size of 100, 200, and 300, respectively; G1 and G2 represent equal correlation G and unequal correlation conditions of G matrix, respectively; gam1 and gam2 represent the same growth pattern across attributes and unequal growth patterns across attributes, respectively.

Since the bias and MSE values of item parameters were not consistent across conditions, ANOVA tests were conducted to examine the impact of design factors on them. When $MO = 3$, resutls found that the sample size had small to large effects on the recoveries on the intercept and main effects parameters ($\eta^2_{\lambda_{0_{\text{Bias}}}} = .05, \eta^2_{\lambda_{\alpha_{\text{Bias}}}} = .15$; $\eta^2_{\lambda_{0_{MSE}}} = .67, \eta^2_{\lambda_{\alpha_{\text{Bias}}}} = .74$). A large sample size was associated with good reroveries. The recoveries of interaction effect parameters were inflenced by the sample size, the G matrix , and the growth pattern. The sample size had large effects on both the bias ($\eta^2_{\lambda_{\alpha_k \alpha_{k'}}}{}_{bias} = .66$) and MSE ($\eta^2_{\lambda_{\alpha_k \alpha_{k'}}}{}_{MSE} = .53$). Similarly, a large sample size resulted in the better recoveries. Both the growth pattern and the G matrix design had small effects on interaction parameter recoveries (the growth pattern: $\eta^2_{\lambda_{\alpha_k \alpha_{k'}}}{}_{bias} = .02, \eta^2_{\lambda_{\alpha_k \alpha_{k'}}}{}_{MSE} = .02$; the G matrix: $\eta^2_{\lambda_{\alpha_k \alpha_{k'}}}{}_{bias} = .02, \eta^2_{\lambda_{\alpha_k \alpha_{k'}}}{}_{MSE} = .02$); the growth and the equal correlations conditions resulted in better recoveries.

When $MO = 5$, the item parameter recoveries were mainly influenced by the sample size. The sample size had small to large effects on the recoveries of intercept and main effects ($\eta^2_{\lambda_{0_{\text{Bias}}}} = .01, \eta^2_{\lambda_{0_{\text{MSE}}}} = .33$; $\eta^2_{\lambda_{\alpha_{\text{Bias}}}} = .05, \eta^2_{\lambda_{\alpha_{\text{MSE}}}} = .37$), and large effects on the recoveries of interaction effects ($\eta^2_{\lambda_{\alpha_k \alpha_{k'}}}{}_{bias} = .19, \eta^2_{\lambda_{\alpha_k \alpha_{k'}}}{}_{MSE} = .17$). The parameter recoveries were improved as the sample size increased. In addition, the recoveries of intercept parameters were influenced by the growth pattern slightly, as shown in Table 4.8. The non-growth condition had a slightly better intercept parameter recoveries, although the effect sizes were very small.

In summary, the item parameter recoveries were mainly influenced by the sample size, especially for the interaction effect parameters. In general, the larger sample size resulted in the better item parameter recoveries.

Table 4.11

ANOVA Results of Measurement Model Parameter Recoveries (MO=3)

Three Measurement Occasions (MO=3)							
		Bias			MSE		
λ_0							
Design factors	*df*	*F*	η^2	*p*	*F*	η^2	*p*
G	1	.	.	.99	2.73	.	.10
gamma	1	.45	.	.50	1.70	.	.19
SZ	2	29.46	.05	.	1156.49	.67	.
G×gamma	1	.07	.	.79	.10	.	.75
G×SZ	2	.21	.	.81	.19	.	.83
gamma×SZ	2	.62	.	.54	2.16	.	.12
G×gamma×SZ	2	1.95	.	.14	.27	.	.77
Residuals	1122		.50			.50	
λ_{α_k}							
Design factors	*df*	*F*	η^2	*p*	*F*	η^2	*p*
G	1	1.13	.	.29	.01	.	.92
gamma	1	1.94	.	.16	1.14	.	.29
SZ	2	96.38	.15	.	1641.27	.75	.
G×gamma	1	.02	.	.90	.11	.	.74
G×SZ	2	.10	.	.90	.16	.	.86
gamma×SZ	2	1.89	.	.15	.82	.	.44
G×gamma×SZ	2	.35	.	.71	1.54	.	.21
Residuals	1122		.50			.50	
$\lambda_{\alpha_k \alpha_{k'}}$							
Design factors	*df*	*F*	η^2	*p*	*F*	η^2	*p*
G	1	27.97	.02	.	24.80	.02	.
gamma	1	23.81	.02	.	21.05	.02	.
SZ	2	1078.32	.66	.	644.66	.53	.
G×gamma	1	.57	.	.45	.19	.	.66
G×SZ	2	.70	.	.49	.06	.	.94
gamma×SZ	2	.72	.	.49	1.14	.	.32
G×gamma×SZ	2	.47	.	.63	.48	.	.62
Residuals	1122		.50			.50	

Note. G represents G matrix design; gamma represents the growth patterns; SZ represents the sample size; . represents < .001.

Table 4.12

ANOVA Results of Measurement Model Parameter Recoveries (MO=5)

Five Measurement Occasions (MO=5)							
		Bias			MSE		
λ_0							
Design factors	*df*	*F*	η^2	*p*	*F*	η^2	*p*
G	1	6.25	.01	.01	5.87	.01	.02
gamma	1	.15	.	.70	7.66	.01	.01
SZ	1	4.96	.01	.03	384.50	.33	.
G×gamma	1	.01	.	.93	2.17	.	.14
G×SZ	1	5.82	.01	.02	.21	.	.65
gamma×SZ	1	1.78	.	.18	1.12	.	.29
G×gamma×SZ	1	4.24	.01	.04	4.70	.01	.03
Residuals	766		.50			.50	
λ_{α_k}							
Design factors	df	F	η^2	p	F	η^2	p
G	1	1.33	.	.25	.33	.	.57
gamma	1	2.20	.	.14	1.68	.	.20
SZ	1	38.92	.05	.	454.47	.37	.
G×gamma	1	.09	.	.77	1.37	.	.24
G×SZ	1	7.44	.01	.01	.17	.	.68
gamma×SZ	1	.37	.	.55	.10	.	.76
G×gamma×SZ	1	2.90	.	.09	2.20	.	.14
Residuals	766		.50			.50	
$\lambda_{\alpha_k \alpha_{k'}}$							
Design factors	df	F	η^2	p	F	η^2	p
G	1	6.75	.01	.01	5.93	.01	.02
gamma	1	.60	.	.44	4.70	.01	.03
SZ	1	176.94	.19	.	158.16	.17	.
G×gamma	1	4.86	.01	.03	2.27	.	.13
G×SZ	1	.85	.	.36	.82	.	.37
gamma×SZ	1	.81	.	.37	2.27	.	.13
G×gamma×SZ	1	2.	.	.16	.02	.	.89
Residuals	766		.50			.50	

Note. G represents G matrix design; gamma represents the growth patterns; SZ represents the sample size; . represents < .001

Structural model parameter recovery.

Recoveries of both fixed effects and random effects in the growth model were evaluated in this study. The fixed effects included the intercept and slope parameters for each attribute ($\gamma_{00}^{A_k}, \gamma_{01}^{A_k}$), and the random effects included the variance of intercept and slope parameters for each attribute ($\delta_{u_0^{A_k}}, \delta_{u_1^{A_k}}$) as well as the covariance among intercept and slope parameters ($\delta_{u_0^{A_k}, u_0^{A_{k'}}}, \delta_{u_1^{A_k}, u_1^{A_{k'}}}, \delta_{u_0^{A_k}, u_1^{A_{k'}}}$).

Recoveries of the fixed effects.

Table 4.13 presents the summary of average bias and MSE of fixed effects under all conditions when $MO = 3$, which reveals that the proposed model achieved good recoveries on the intercept parameters for Attributes 2 and 3, and slope parameters for all attributes, indicated by the small MSE values and the bias values being closed to zero. However, the intercept parameter of Attribute 1 had relatively larger bias than other parameters.

The bias and MSE of intercept parameters were not consistent across different conditions, so ANOVA tests were conducted to investigate if the design factors influenced the intercept parameter recoveries. As shown in Table 4.15, that the sample size had small effects on the MSE values of intercept parameters (${\eta^2}_{\gamma_{00}}^{A1} = .03, {\eta^2}_{\gamma_{00}}^{A2} = .04, {\eta^2}_{\gamma_{00}}^{A3} = .03$). A large sample size was associated with small MSE values. However, the bias of fixed effects were not influenced by the design factors.

When $MO = 5$, similar patterns were found as shown in Table 4.14. The intercept parameters of Attributes 2 and 3, and the slope parameters for all attributes were recoveried well, but the intercept of Attribute 1 had relatively larger bias than other parameters.

As shown in Tables 4.16, ANOVA tests found that the sample size had small effects on the MSE values of intercept parameters for Attribute 2 and 3 (${\eta^2}^{A2}_{\gamma_{00}} = .01, {\eta^2}^{A3}_{\gamma_{00}} = .01$). Similarly, the bias of intercept parameters were not influenced by the design factors.

In summary, the intercept parameters of Attributes 2 and 3 and all the slope parameters were recovered well in the current study, but the intercept parameters of Attribute 1 had a relatively large bias. ANOVA tests found that the sample size had small effects on the MSE values of intercept parameters; a larger sample size resulted in smaller MSE values. However, no design factors were associated with the bias of intercept parameters.

Table 4.13

Summary of Fixed Effects Recoveries (MO=3)

			γ_{00}^{A1}		γ_{01}^{A1}		γ_{00}^{A2}		γ_{01}^{A2}		γ_{00}^{A3}		γ_{01}^{A3}	
			Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
G1	gam1	N100	-.10	.11	.	.01	-.02	.08	.	.01	.07	.10	-.02	.01
		N200	-.11	.08	-.01	.01	.01	.07	.	.01	.03	.08	.	.01
		N300	-.11	.05	.	.	.02	.06	.	.01	.04	.06	-.01	.
	gam2	N100	-.11	.11	.	.01	.09	.11	.02	.01	.06	.10	.	.01
		N200	-.13	.07	.01	.01	-.01	.06	.02	.01	.06	.07	.	.01
		N300	-.10	.06	-.01	.	.02	.04	.	.	.05	.05	.	.
G2	gam1	N100	-.12	.09	-.01	.01	.	.09	.	.01	.02	.08	.	.01
		N200	-.09	.08	.01	.01	-.01	.05	.	.01	.05	.07	.01	.01
		N300	-.11	.06	.01	.	.02	.05	.01	.01	.02	.06	.	.
	gam2	N100	-.11	.10	-.01	.01	.03	.09	.02	.01	.	.14	.	.01
		N200	-.11	.06	.	.01	.	.06	.	.01	.02	.05	.02	.01
		N300	-.11	.07	.01	.	-.01	.04	.	.	.05	.05	.	.01

Note. γ_{00}^{k} and γ_{01}^{k} represents the intercept and slope parameters of attributes; A1, A2, and A3 represent Attribute 1, Attribute 2, and Attribute 3; N100, N200, and N300 represent the sample size of 100, 200, and 300, respectively; . represents < .001.

Table 4.14

Summary of Fixed Effects Recoveries (MO=5)

			γ_{00}^{A1}		γ_{01}^{A1}		γ_{00}^{A2}		γ_{01}^{A2}		γ_{00}^{A3}		γ_{01}^{A3}	
			Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
G1	gam1	N100	-.11	.08	.	.01	.01	.04	.	.01	.02	.06	.	
		N200	-.13	.06	-.01	.	.02	.05	.	.	.05	.05	.	.
		N300	-.10	.04	.	.	-.01	.03	.01	.	.03	.04	-.01	.
	gam2	N100	-.12	.08	.01	.01	.01	.06	.	.01	.01	.05	.	.01
		N200	-.13	.06	-.02	.01	.02	.05	.01	.	.05	.06	.01	.01
		N300	-.15	.06	-.01	.	.	.04	.01	.	.10	.04	-.01	.
G2	gam1	N100	-.14	.09	-.01	.01	-.01	.07	.	.01	.07	.08	.01	.01
		N200	-.16	.06	.01	.01	.01	.04	.	.	.03	.04	-.01	.01
		N300	-.11	.06	.01	.01	-.02	.04	.01	.	.01	.03	-.01	.
	gam2	N100	-.12	.07	-.01	.01	-.03	.07	-.01	.01	.04	.07	.01	.01
		N200	-.13	.07	.01	.	-.04	.05	.	.	.06	.05	.	.
		N300	-.11	.05	.	.	-.02	.03	.01	.	.06	.04	.	.

Note. γ_{00}^{k} and γ_{01}^{k} represents the intercept and slope parameters of attributes; A1, A2, and A3 represent Attribute 1, Attribute 2, and Attribute 3; N100, N200, and N300 represent the sample size of 100, 200, and 300, respectively; G1 and G2 represent equal correlation G and unequal correlation conditions of G matrix, respectively; gam1 and gam2 represent the same growth pattern across attributes and unequal growth patterns across attributes, respectively; . represents < .001.

Table 4.15

ANOVA Results of Fixed Effects Parameter Recoveries (MO=3)

	Three Measurement Occasions (MO=3)						
		Bias			RMSE		
Design factors	df	F	η^2	p	F	η^2	p
γ_{00}^{A1}							
G	1	.10	.	.75	.37	.	.54
SZ	2	.06	.	.94	14.93	.03	.
G×SZ	2	.42	.	.66	.83	.	.43
Residuals	1128		.50			.50	
γ_{00}^{A2}							
G	1	.89	.	.35	2.78	.	.10
SZ	2	.72	.	.49	22.14	.04	.
G×SZ	2	.05	.	.95	.23	.	.79
Residuals	1128		.50			.50	
γ_{00}^{A3}							
G	1	2.31	.	.13	.28	.	.60
SZ	2	.03	.	.97	15.78	.03	.
G×SZ	2	1.07	.	.34	1.46	.	.23
Residuals	1128		.50			.50	

Note. G represents G matrix design; gamma represents the growth patterns; SZ represents the sample size; . represents < .001.

Table 4.16

ANOVA Results of Fixed Effects Parameter Recoveries (MO=5)

Five Measurement Occasions (MO=5)							
		Bias			RMSE		
Design factors	df	F	η^2	p	F	η^2	p
γ_{00}^{A1}							
G	1	.05	.	.82	1.05	.	.31
SZ	1	2.04	.	.15	1.11	.	.29
G×SZ	1	1.19	.	.28	.11	.	.74
Residuals	770		.50			.50	
γ_{00}^{A2}							
G	1	2.57	.	.11	.37	.	.55
SZ	1	1.02	.	.31	8.58	.01	.
G×SZ	1	.22	.	.64	.08	.	.78
Residuals	770		.50			.50	
γ_{00}^{A3}							
G	1	1.26	.	.26	.42	.	.52
SZ	1	.20	.	.65	6.65	.01	.01
G×SZ	1	.38	.	.54	1.40	.	.24
Residuals	770		.50			.50	

Note. G represents G matrix design; gamma represents the growth patterns; SZ represents the sample size; . represents < .001.

Recoveries of random effects.

Recovery of variance parameters. Table 4.17 presents the average bias and MSE values of the variance of intercept and slope for all attributes when $MO = 3$, which reveals that the proposed model achieved good recoveries in both the intercept and slope variance parameters. Since bias of intercept variance parameters were not consistent across all conditions, ANOVA tests were conducted to examine the impact of design factors on them. As shown in Table 4.19, results found that the sample size had medium effects on the bias of intercept variance parameters ($\eta^2_{\delta_{u_0^{A_1}}} = .14$; $\eta^2_{\delta_{u_0^{A_2}}} = .13$; $\eta^2_{\delta_{u_0^{A_3}}} = .11$); the large sample size had large bias values.

When $MO = 5$, similar patterns were found. The variance of intercept and slope parameters were recovered well. Since the recoveries of the varaince of intercept parameters were varied by conditions, ANOVA tests were conducted to investigate the impact of design factors on them. As showed in Table 4.20, the sample size had small effects ($\eta^2_{\delta_{u_0^{A_1}}} = .02$; $\eta^2_{\delta_{u_0^{A_2}}} = .02$; $\eta^2_{\delta_{u_0^{A_3}}} = .02$); the larger sample size had larger bias values.

In summary, the proposed model achieved good recoveries on the variance of intercept and slope parameters. Moreover, a large sample size was associated with large bias values of the variance of intercept parameters.

Table 4.17

Summary of Random Variance Recoveries (MO=3)

			$\delta_{u_0^{A_1}}$		$\delta_{u_0^{A_2}}$		$\delta_{u_0^{A_3}}$		$\delta_{u_1^{A_1}}$		$\delta_{u_1^{A_2}}$		$\delta_{u_1^{A_3}}$	
			Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
G1	gam1	N100	.02	.01	.03	.01	.02	.01	.	.	.	.	.01	.
		N200	-.01	.	-.01	.	.	.	.	.	.	.	.	.
		N300	-.03	.	-.03	.	-.03	.	.	.	.	.	.	.
	gam2	N100	.02	.	.02	.	.01	.	.01	.	.01	.	.01	.
		N200	-.02	.	-.01	.	-.01	.	.	.	.	.	.	.
		N300	-.03	.	-.03	.	-.03	.	.	.	.	.	.	.
G2	gam1	N100	.04	.01	.04	.01	.04	.01	.01	.	.01	.	.01	.
		N200	-.01	.	-.01	.	-.01	.01	.	.	.	.	.	.
		N300	-.04	.	-.03	.	-.03	.	.	.	.	.	.	.
	gam2	N100	.03	.01	.03	.01	.03	.01	.01	.	.01	.	.	.
		N200	-.02	.	-.02	.	-.02	.	.01	.	.	.	.	.
		N300	-.03	.	-.03	.	-.03	.	.	.	.	.	.	.

Note: A1, A2, and A3 represent Attribute 1, Attribute 2, and Attribute 3; N100, N200, and N300 represent the sample size of 100, 200, and 300, respectively; G1 and G2 represent equal correlation G and unequal correlation conditions of G matrix, respectively; gam1 and gam2 represent the same growth pattern across attributes and unequal growth patterns across attributes, respectively; . represents < .001.

Table 4.18

Summary of Random Variance Recoveries (MO=5)

			$\delta_{u_0^{A_1}}$		$\delta_{u_0^{A_2}}$		$\delta_{u_0^{A_3}}$		$\delta_{u_1^{A_1}}$		$\delta_{u_1^{A_2}}$		$\delta_{u_1^{A_3}}$	
			Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
G1	gam1	N100	.	.	.	.	-.01	.	.01	.	.01	.	.	.
		N200	-.03	.	-.03	.	-.03	.	.	.	.	.	.	.
		N300	-.04	.	-.04	.	-.04	.	.	.	.	.	.	.
	gam2	N100	.01	.01	.01	.	.	.	.	.	.	.	.	.
		N200	-.03	.	-.03	.	-.04	.	.	.	.	.	.	.
		N300	-.04	.	-.04	.	-.04	.	.	.	.	.	.	.
G2	gam1	N100	.	.	.	.	.	.	.01	.	.	.	.01	.
		N200	-.04	.	-.03	.	-.04	.	.	.	.	.	.	.
		N300	-.05	.	-.04	.	-.04	.	.	.	.	.	.	.
	gam2	N100	.	.	-.01	.	-.01	.	.01	.	.	.	.01	.
		N200	-.03	.	-.03	.	-.03	.	.	.	.	.	.	.
		N300	-.05	.	-.05	.	-.05	.	.	.	.	.	.	.

Note: A1, A2, and A3 represent Attribute 1, Attribute 2, and Attribute 3; N100, N200, and N300 represent the sample size of 100, 200, and 300, respectively; G1 and G2 represent equal correlation G and unequal correlation conditions of G matrix, respectively; gam1 and gam2 represent the same growth pattern across attributes and unequal growth patterns across attributes, respectively; . represents < .001.

Table 4.19

ANOVA Results of Random Variance Parameter Recoveries (MO=3)

Three Measurement Occasions (MO=3)							
		Bias			RMSE		
Design factors	*df*	*F*	η^2	*p*	*F*	η^2	*p*
γ_{00}^{A1}							
G	1	.10	.	.75	1.53	.	.22
SZ	2	89.68	.14	.	1.27	.02	.
G×SZ	2	2.58	.	.08	1.12	.	.33
Residuals	1128		.50			.50	
γ_{00}^{A2}							
G	1	.08	.	.78	2.79	.	.10
SZ	2	84.06	.13	.	11.79	.02	.
G×SZ	2	1.78	.	.17	2.19	.	.11
Residuals	1128		.50			.50	
γ_{00}^{A3}							
G	1	.14	.	.71	3.54	.	.06
SZ	2	72.67	.11	.	7.26	.01	.
G×SZ	2	3.01	.01	.05	2.37	.	.09
Residuals	1128		.50			.50	

Note. G represents G matrix design; gamma represents the growth patterns; SZ represents the sample size; . represents < .001.

Table 4.20 ANOVA Results of Random Variance Parameter Recoveries (MO=5)

ANOVA Results of Random Variance Parameter Recoveries (MO=5)

Five Measurement Occasions (MO=5)							
		Bias			RMSE		
Design factors	*df*	*F*	η^2	*p*	*F*	η^2	*p*
γ_{00}^{A1}							
G	1	2.21	.	.14	.	.	.97
SZ	1	14.49	.02	.	2.35	.	.13
G×SZ	1	.01	.	.94	.19	.	.66
Residuals	770		.50			.50	
γ_{00}^{A2}							
G	1	.	.	.97	.54	.	.46
SZ	1	13.35	.02	.	4.22	.01	.04
G×SZ	1	.01	.	.92	1.48	.	.22
Residuals	770		.50			.50	
γ_{00}^{A3}							
G	1	.64	.	.43	.	.	1.
SZ	1	14.08	.02	.	6.48	.01	.01
G×SZ	1	.08	.	.77	.59	.	.44
Residuals	770		.50			.50	

Note. G represents G matrix design; gamma represents the growth patterns; SZ represents the sample size; . represents < .001.

Recovery of covariance parameters. As shown in Table 4.17, on average, the proposed model achieved good recoveries on the covariance among intercept and slope parameters. However, the covariance between intercepts had lightly larger bias than other sets of parameters. Then, ANOVA tests found that the sample size had medium effects ($\eta^2 = .13$) on the bias of convariance between intercept parameters; a large sample size was associated with a large bias.

Similar patterns were found when $MO = 5$. The proposed model achieved good recoveries on the covariance between intercept and slope parameters. Similarly, the covariance between intercept parameters had slightly larger bias than other sets of parameters. Then, ANOVA tests found the sample size had medium effects on the bias values of covariance between intercept parameters; a larger sample size was associated with a larger bias value.

On average, the proposed model achieved the good recoveries on the covariance among intercept and slope parameters. The bias of covariance among intercept parameters was influenced by the sample size; the larger sample size resulted in larger bias values.

Table 4.21

Summary of Covariance Parameter Recoveries

			Three Measurement Occasions (MO=3)						Five Measurement Occasions (MO=5)					
			$\delta_{u_0^{A_k}, u_0^{A_{k'}}}$		$\delta_{u_1^{A_k}, u_1^{A_{k'}}}$		$\delta_{u_0^{A_k}, u_1^{A_{k'}}}$		$\delta_{u_0^{A_k}, u_0^{A_{k'}}}$		$\delta_{u_1^{A_k}, u_1^{A_{k'}}}$		$\delta_{u_0^{A_k}, u_1^{A_{k'}}}$	
			Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
G1	gam1	N100	.02	.01	.	.	.	.	-.01	.	.	.	.	.
		N200	-.01	.	.	.	.	.	-.04	.	.	.	.	.
		N300	-.03	.	.	.	.	.	-.05	.	.	.	.	.
	gam2	N100	.01	.	.	.	.	.	.	.	.	.	.	.
		N200	-.02	.	.	.	.	.	-.04	.	.	.	.	.
		N300	-.03	.	.	.	.	.	-.05	.	.	.	.	.
G2	gam1	N100	.03	.01	.	.	.	.	-.01	.	.	.	.	.
		N200	-.02	.	.	.	.	.	-.04	.	.	.	.	.
		N300	-.04	.	.	.	.	.	-.05	.	.	.	.01	.
	gam2	N100	.02	.01	.	.	.	.	-.02	.	.	.	.	.
		N200	-.03	.	.	.	.	.	-.04	.	.	.	.	.
		N300	-.03	.	.	.	.	.	-.05	.	.		.	

Note: A1, A2, and A3 represent Attribute 1, Attribute 2, and Attribute 3; N100, N200, and N300 represent the sample size of 100, 200, and 300, respectively; G1 and G2 represent equal correlation G and unequal correlation conditions of G matrix, respectively; gam1 and gam2 represent the same growth pattern across attributes and unequal growth patterns across attributes, respectively; . represents < .001.

Table 4.22

ANOVA Results of Random Covariance Parameter Recoveries (MO=3)

Three Measurement Occasions (MO=3)							
		Bias			MSE		
Design factors	*df*	*F*	η^2	*p*	*F*	η^2	*p*
$\delta_{u_0^{A_k}, u_0^{A_{k'}}}$							
G	1	.16	.	.69	3.20	.	.07
SZ	2	81.22	.13	.	6.96	.01	.
G×SZ	2	2.87	.01	.06	2.04	.	.13
Residuals	1128		.50			.50	

Note. G represents G matrix design; gamma represents the growth patterns; SZ represents the sample size; . represents < .001.

Table 4.23

ANOVA Results of Random Covariance Parameter Recoveries (MO=5)

Five Measurement Occasions (MO=5)							
		Bias			MSE		
Design factors	*df*	*F*	η^2	*p*	*F*	η^2	*p*
$\delta_{u_0^{A_k}, u_0^{A_{k'}}}$							
G	1	.82	.	.37	.10	.	.75
SZ	1	5.32	.01	.02	5.81	.01	.02
G×SZ	1	.25	.	.62	.70	.	.40
Residuals	770		.50			.50	

Note. G represents G matrix design; gamma represents the growth patterns; SZ represents the sample size, . represents < .001.

Chapter 5 : Discussion

This dissertation proposed a multivariate longitudinal DCM that could incorporate balanced data as well as unbalanced data and measure the growth of multiple attributes directly without assuming they have similar growth trajectories. The primary goal of this dissertation was to evaluate the proposed model under several conditions. Specifically, it sought to answer the following research questions:

Does the proposed model provide satisfied classification accuracy under different conditions?

Do the sample size, the growth patterns, the G matrix design, the number of measurement occasions, and their interactions impact the parameter recoveries in the measurement model?

Do the sample size, the growth patterns, the G matrix design, the number of measurement occasions, and their interactions impact the parameter recoveries in the growth model?

To address the research questions listed above, one simulation study was conducted. This simulation study evaluated the proposed model in terms of the accuracy of classification and parameter recoveries under different combinations of four design factors: the sample size, the growth patterns, the G matrix design, and the number of measurement occasions.

This chapter includes three sections. It begins with a summary and discussion of the simulation study. Then, it provides a general conclusion and recommendations for applied researchers. Lastly, it concludes by discussing the contributions and limitations of the current study.

Performance of the Multivariate Longitudinal DCM

Model convergence.

Overall, the proposed model achieved satisfied convergence rates; however, the proposed achieved a slightly higher convergence rates when $MO = 5$ than $MO = 3$, which was reasonable since more measurement occasions would provide more information to help the estimation and the model be converged. Also, as shown in Appendix B, the conditions with five measurement occasions had more chains and a longer chain length for each chain than the conditions with three measurement occasions, which might have led to improve the convergence rates. Therefore, the number of chains and the chain length might be not sufficient for the conditions with three measurement occasions.

Classification accuracy.

The bias of the estimated probability of attribute mastery, the correct classification rates for each mastery status, and Cohen's kappa was used to evaluate the classification accuracy of the proposed model.

The probability of attribute mastery was recovered well in the current study consistently across all measurement occasion, which indicated that the proposed model could provide accurate estimates of probabilities of attribute mastery.

Regarding correct classification rates, results found different patterns for individuals who truly mastered the attributes and individuals who truly did not master the attributes. For the individuals who truly mastered the attributes, the correct classification rates improved significantly as the number of measurement occasions increased. However, for individuals who truly did not master the attributes, the correct classification rates decreased slightly as the number of measurement occasions increased. This pattern might be due to the cutpoint that was adopted to classify the individuals. Since the estimated probabilities of attribute mastery for the most of individuals were lower than 0.5 at the first two measurement occasions, individuals would be classified into the non-mastery status, even some of them truly mastered the attributes by design. As a result, the correct classification rates were low for individuals who truly mastered the attributes at the first two measurement occasions. As the number of measurement occasions increased, the estimated probability of mastery increased, such that correct classification rates increased. Due to the same reason, Cohen's kappa increased as the number of measurement occasions increased. Therefore, the cutpoint had influenced the correct classification rates and kappa values of the current model.

Parameter recoveries.

The bias and mean square error (MSE) of the estimated parameters were computed to assess the parameter recovery in each condition through the simulation.

Measurement model parameter recoveries. Regarding the item parameter recoveries, conditions with three and five measurement occasions illustrated similar patterns. The proposed model achieved good parameter recoveries in intercept and main effect parameters, but poor interaction effect parameter recoveries. However, the recoveries of the interaction effect parameters were improved as the sample size and the number of measurement occasions increased. In addition, results from the ANOVA tests found the sample size had large impact on the interaction effects recoveries. Nonetheless, this result was expected. Previous research showed that the intercept and main effect parameters were easier to recover than the two-way interaction effect parameters. The recoveries of the interaction effect parameters were problematic when the sample size was less than 1000 (e.g., Choi, Templin, Cohen, & Atwood, 2010; Kunina-Habenicht et al., 2012). Therefore, these results suggested that a large sample size was necessary to achieve good item parameter recoveries in the LCDM framework, especially for the interaction effect parameters. The maximum sample size (N=300) in the current study was not sufficient for obtaining accurate interaction effect parameters, especially for the conditions with three measurement occasions.

Structural model parameter recoveries. Both the recoveries of fixed effects and random effects in the generalized growth curve model were evaluated.

Regarding the recoveries of the fixed effects, overall, the proposed model achieved good intercept recoveries for Attributes 2 and 3, and slope recoveries for all attributes, but relatively poor recoveries for Attribute 1 intercept. The Attribute 1 had relatively small intercept value by design ($\gamma_{00}^{A1} = -1.38$), therefore, the small intercept value might have led to enlarge the bias. To avoide the influence of the small value of the intercept parameter, the time variable could be centered at the medial measurement occasions ($T = 2$ when $MO = 3$, or $T = 3$ when $MO = 5$), such that there would be sufficient information to estimate the intercept parameters.

Regarding the recoveries of the random effects, on average, the proposed model achieved good recoveries on the random effects, including the variance of intercept and slope parameters of each attribute as well as the covariance among intercept and slope parameters within and crossed attributes. To improve the model convergence, the current study adopted the true variance-covariance matrix in the population as the prior of the estimated variance-covariance matrix, which might have led to good recoveries of the random effects.

Conclusion and Recommendations

The current study developed a multivariate longitudinal DCM that could measure growth in attributes over time, and it evaluated this proposed model using a simulation study. The results revealed the following: (1) In general, the proposed model provided good convergence rates under different conditions. (2) Regarding the classification accuracy, the proposed model achieved good recoveries on the probabilities of attribute mastery. For individuals who truly mastered the attributes, the correct classification rates increased as the measurement occasions increased; however, for individuals who truly did not master the attributes, the correct classification rates decreased slightly as the numbers of measurement occasions increased. Cohen's kappa increased as the number of measurement occasions increased. (3) Both the intercept and main effect parameters in the LCDM were recovered well. The interaction effect parameters had a relatively large bias under the condition with small sample size and fewer measurement occasions; however, the recoveries were improved as the sample size and the number of measurement occasions increased. (4) Overall, the proposed model achieved acceptable recoveries on both the fixed and random effects in the generalized growth curve model.

In summary, a large sample size is recommended for applying the proposed model to the real data. When the sample size is small, the scale with a simple structure of Q matrix is recommended, because the interaction effects in the LCDM might not be estimated accurately with the small sample size. Also, applied researchers are suggested to center the time variable at the medial measurement occasion to improve the recovery of the intercept parameter in the generalized growth curve model. Additionally, when doing the MCMC analysis, multiple chains with the longer chain length are recommended to achieve satisfied model convergence rates.

Therefore, when practitioners try to measure students' growth in the DCM framework using the proposed model, they should use a larger sample size, an assessment with less complex Q-matrix design, and multiple chains with longer chain length to maximize the convergence rates and the accuracy of parameter estimates.

Contributions and Limitations

In the current study, a multivariate longitudinal DCM was developed to analyze longitudinal data under the DCM framework. It represents an improvement in the current longitudinal DCMs given its ability to incorporate both balanced and unbalanced data and to measure the growth of a single attribute directly without assuming that attributes grow in the same pattern.

The current study had several limitations. First, the true variance-covariance matrix was used as the prior for the random effects parameters in the generalized growth curve model in the current study; however, the true variance-covariance matrix is unknown when fitting the model to the real data. Therefore, future studies could adopted non-informative variance-covariance matrix as the prior, then evaluate if the proposed model could achieve satisfied recoveries on the random effects as well. Second, local item dependency was not incorporated in the current study. However, in real longitudinal data, repeated measures always have some degree of local item dependency (e.g., Cai, 2010). Therefore, future research could simulate local item dependence with the common items to mimic real data. Third, only three or five measurement occasions were included in the current model. The small number of measurement occasions might have limited the reliability and accuracy of the estimation of the growth curve model (e.g., Finch, 2017). In the future, more measurement occasions could be included to examine the performance of the proposed model comprehensively.

Reference

Bradshaw, L., Izsák, A., Templin, J., & Jacobson, E. (2014). Diagnosing teachers' understandings of rational numbers: Building a multidimensional test within the diagnostic classification framework. *Educational measurement: Issues and practice, 33*(1), 2-14.

Bradshaw, L., & Madison, M. J. (2016). Invariance Properties for General Diagnostic Classification Models. *International Journal of Testing, 16*(2), 99-118. doi:10.1080/15305058.2015.1107076

Bradshaw, L., & Templin, J. (2014). Combining item response theory and diagnostic classification models: A psychometric model for scaling ability and diagnosing misconceptions. *Psychometrika, 79*(3), 403-425.

Brown, T. A. (2014). *Confirmatory factor analysis for applied research*: Guilford Publications.

Cai, L. (2010). A Two-Tier Full-Information Item Factor Analysis Model with Applications. *Psychometrika, 75*(4), 581-612. doi:10.1007/s11336-010-9178-0

Chen, J., & Torre, J. d. l. (2013). A General Cognitive Diagnosis Model for Expert-Defined Polytomous Attributes. *Applied Psychological Measurement, 37*(6), 419-437. doi:10.1177/0146621613479818

Choi, H.-J., Templin, J., Cohen, A., & Atwood, C. (2010). *The impact of model misspecification on estimation accuracy in diagnostic classification models*. Paper presented at the National Council on Measurement in Education (NCME), Denver, CO.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement, 20*(1), 37-46. doi:10.1177/001316446002000104

Cohen, J. (1988). Statistical power analysis for the behavioral sciences . Hilsdale. *NJ: Lawrence Earlbaum Associates, 2*.

Collins, L. M., & Wugalter, S. E. (1992). Latent Class Models for Stage-Sequential Dynamic Latent Variables. *Multivariate Behavioral Research, 27*(1), 131-157. doi:10.1207/s15327906mbr2701_8

Curran, P. J., McGinley, J. S., Serrano, D., & Burfeind, C. (2012). A multivariate growth curve model for three-level data. In H. Cooper (Ed.), *APA Handbook of Research Methods in Psychology* (Vol. 3, pp. 335-358). Washington, DC: American Psychological Association.

Davier, M. (2005). A general diagnostic model applied to language testing data. *ETS Research Report Series, 2005*(2).

de La Torre, J. (2008). An Empirically Based Method of Q-Matrix Validation for the DINA Model: Development and Applications. *Journal of Educational Measurement, 45*(4), 343-362. doi:10.1111/j.1745-3984.2008.00069.x

de La Torre, J. (2009). A cognitive diagnosis model for cognitively based multiple-choice options. *Applied Psychological Measurement, 33*(3), 163-183.

de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika, 69*(3), 333-353. doi:10.1007/bf02295640

de La Torre, J., Hong, Y., & Deng, W. (2010). Factors Affecting the Item Parameter Estimation and Classification Accuracy of the DINA Model. *Journal of Educational Measurement, 47*(2), 227-249. doi:10.1111/j.1745-3984.2010.00110.x

de la Torre, J. (2011). The Generalized DINA Model Framework. *Psychometrika, 76*(2), 179-199. doi:10.1007/s11336-011-9207-7

DiBello, L. V., Stout, W. F., & Roussos, L. A. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. *Cognitively diagnostic assessment*, 361-389.

Dynamic Learning Maps. (2016). *2014–2015 Technical Manual: Integrated Model*. Retrieved from Lawrence, KS, USA:

Finch, W. H. (2017). Investigation of Parameter Estimation Accuracy for Growth Curve Modeling With Categorical Indicators. *Methodology, 13*(3), 98-112. doi:10.1027/1614-2241/a000134

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science, 7*(4), 457-472.

George, A. C., & Robitzsch, A. (2014). Multiple group cognitive diagnosis models, with an emphasis on differential item functioning. *Psychological Test and Assessment Modeling, 56*(4), 405-432.

Goldstein, H. (2011). *Multilevel statistical models* (Vol. 922). New York, NY: Halsted: John Wiley & Sons.

Goodman, L. A. (1974). Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models. *Biometrika, 61*(2), 215-231. doi:10.2307/2334349

Great Schools Partnership. (2013, 10.30.2013). Interim Assessment. Retrieved from http://edglossary.org/interim-assessment/

Hartz, S. M. (2002). *A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality.* University of Illinois at Urbana-Champaign,

Hoffman, L. (2015). *Longitudinal analysis: Modeling within-person fluctuation and change*. New York, NY: Routledge.

Huang, H.-Y. (2017). Multilevel Cognitive Diagnosis Models for Assessing Changes in Latent Attributes. *Journal of Educational Measurement, 54*(4), 440-480. doi:10.1111/jedm.12156

Junker, B. W., & Sijtsma, K. (2001). Cognitive Assessment Models with Few Assumptions, and Connections with Nonparametric Item Response Theory. *Applied Psychological Measurement, 25*(3), 258-272. doi:10.1177/01466210122032064

Kaya, Y., & Leite, W. L. (2017). Assessing change in latent skills across time with longitudinal cognitive diagnosis modeling: An evaluation of model performance. *Educational and Psychological Measurement, 77*(3), 369-388.

Kunina-Habenicht, O., Rupp, A. A., & Wilhelm, O. (2012). The Impact of Model Misspecification on Parameter Estimation and Item-Fit Assessment in Log-Linear Diagnostic Classification Models. *Journal of Educational Measurement, 49*(1), 59-81. doi:10.1111/j.1745-3984.2011.00160.x

Lanza, S. T., Flaherty, B. P., & Collins, L. M. (2003). Latent class and latent transition analysis. *Handbook of psychology*.

Lazarsfeld, P., & Henry, N. (1968). Latent Structure Analysis. In: Houghton Mifflin Company, Boston, Massachusetts.

Lee, Y.-W., & Sawaki, Y. (2009). Application of three cognitive diagnosis models to ESL reading and listening assessments. *Language Assessment Quarterly, 6*(3), 239-263.

Leighton, J., & Gierl, M. (2007). *Cognitive diagnostic assessment for education: Theory and applications*: Cambridge University Press.

Li, F., Cohen, A., Bottge, B., & Templin, J. (2016). A latent transition analysis model for assessing change in cognitive skills. *Educational and Psychological Measurement, 76*(2), 181-204.

MacCallum, R. C., Kim, C., Malarkey, W. B., & Kiecolt-Glaser, J. K. (1997). Studying Multivariate Change Using Multilevel Models and Latent Curve Models. *Multivariate Behavioral Research, 32*(3), 215-253. doi:10.1207/s15327906mbr3203_1

Madison, M. J. (2016). *Analyzing pre-test/post-test design in diagnostic classification model framework.* (Ph.D.), The University of Georgia.,

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika, 64*(2), 187-212. doi:10.1007/bf02294535

Preacher, K. J., Wichman, A., MacCallum, R. C., & Briggs, N. E. (2008). *Latent growth curve modeling* (V. Knight Ed.). thousand Oaks, CA: Sage.

R Core Team. (2017). R: A language and environment for statistical computing (Version 3.4.2): R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Raghavarao, D., & Padgett, L. (2014). Growth Curve Models. In *Repeated Measurements and Cross-Over Designs* (pp. 77-104). Hoboken, NJ: John Wiley & Sons, Inc.

Ravand, H. (2016). Application of a cognitive diagnostic model to a high-stakes reading comprehension test. *Journal of Psychoeducational Assessment, 34*(8), 782-799.

Roussos, L. A., DiBello, L. V., Stout, W., Hartz, S. M., Henson, R. A., & Templin, J. (2007). The fusion model skills diagnosis system. *Cognitive diagnostic assessment for education: Theory and applications*, 275-318.

Rupp, A. A., & Templin, J. (2008a). The Effects of Q-Matrix Misspecification on Parameter Estimates and Classification Accuracy in the DINA Model. *Educational and Psychological Measurement, 68*(1), 78-96. doi:10.1177/0013164407301545

Rupp, A. A., & Templin, J. (2008b). Unique Characteristics of Diagnostic Classification Models: A Comprehensive Review of the Current State-of-the-Art. *Measurement: Interdisciplinary Research and Perspectives, 6*(4), 219-262. doi:10.1080/15366360802490866

Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*: Guilford Press.

Sedat, Ş., & Arican, M. (2015). A Diagnostic Comparison of Turkish and Korean Students' Mathematics Performances on the TIMSS 2011 Assessment. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 6*(2).

Sinharay, S. (2003). ASSESSING CONVERGENCE OF THE MARKOV CHAIN MONTE CARLO ALGORITHMS: A REVIEW. *ETS Research Report Series, 2003*(1), i-52. doi:10.1002/j.2333-8504.2003.tb01899.x

Templin, J., & Bradshaw, L. (2014). Hierarchical diagnostic classification models: A family of models for estimating and testing attribute hierarchies. *Psychometrika, 79*(2), 317-339.

Templin, J., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological methods, 11*(3), 287. doi:10.1037/1082-989X.11.3.287

Templin, J., Henson, R. A., Templin, S. E., & Roussos, L. (2008). Robustness of Hierarchical Modeling of Skill Association in Cognitive Diagnosis Models. *Applied Psychological Measurement, 32*(7), 559-574. doi:10.1177/0146621607300286

Torre, J. d. l. (2009). DINA Model and Parameter Estimation: A Didactic. *Journal of Educational and Behavioral Statistics, 34*(1), 115-130. doi:10.3102/1076998607309474

von Davier, M. (2005). A general diagnostic model applied to language testing data. *ETS Research Report Series, 2005*(2), i-35. doi:10.1002/j.2333-8504.2005.tb01993.x

Zhan, P., Jiao, H., & Liao, D. (2017). *A Longitudinal Diagnostic Classification Model*. arXiv preprint arXiv:1709.03431.

Appendix A

Jags code for the multivariate longitudinal DCM

```
model{
# the LCDM
for(t in 1:T){ # time points
for(n in 1:N){ # people
for(i in 1:I){ # item
logit(p[n, i, t]) <- lamda0[i] + lamda1[i] * alpha[n,1,t] * Q[i,1] + lamda2[i] * alpha[n,2,t] * Q[i,2] + lamda3[i] * alpha[n,3,t] * Q[i,3] + lamda12[i] * alpha[n,1,t] * Q[i,1] * alpha[n,2,t] * Q[i,2] + lamda13[i] * alpha[n,1,t] * Q[i,1] * alpha[n,3,t] * Q[i,3] + lamda23[i] * alpha[n,2,t] * Q[i,2] * alpha[n,3,t] * Q[i,3]
Y[n, i, t] ~ dbern(p[n, i, t])}
}}
## growth model
  for(t in 1:T){
      for(n in 1:N){
logit(prob[n,1,t]) <- gamma[1,1]+ u[n,1] + (gamma[1,2] + u[n,2])*PersonTime[n,t]
logit(prob[n,2,t]) <- gamma[2,1]+ u[n,3] + (gamma[2,2] + u[n,4])*PersonTime[n,t]
logit(prob[n,3,t]) <- gamma[3,1]+ u[n,5] + (gamma[3,2] + u[n,6])*PersonTime[n,t]
alpha[n,1,t] ~ dbern(prob[n,1,t])
alpha[n,2,t] ~ dbern(prob[n,2,t])
alpha[n,3,t] ~ dbern(prob[n,3,t])
}}
# prior for the growth model
# for random effects
```

```
for( n in 1:N){
  u[n,1:6] ~ dmnorm(mean.u[1:6],latprec.u[1:6,1:6])
}
for(g in 1:6){
mean.u[g] ~ dnorm(0,.0001)
}

 latprec.u[1:6,1:6] ~ dwish(cov.u,6)
 latcov.u[1:6,1:6] <- inverse(latprec.u[1:6,1:6])

# for fixed effects
gamma[1,1] ~ dnorm(mean.int[1],.001)
gamma[2,1] ~ dnorm(mean.int[2],.001)
gamma[3,1] ~ dnorm(mean.int[3],.001)

gamma[1,2] ~ dnorm(mean.slp[1],.001)
gamma[2,2] ~ dnorm(mean.slp[2],.001)
gamma[3,2] ~ dnorm(mean.slp[3],.001)

mean.int[1] ~ dnorm(0,.001)
mean.int[2] ~ dnorm(0,.001)
mean.int[3] ~ dnorm(0,.001)

mean.slp[1] ~ dnorm(0,.001)
```

```
mean.slp[2] ~ dnorm(0,.001)

mean.slp[3] ~ dnorm(0,.001)

# prior for the LCDM

for(i in 1:I) {

lamda0[i] ~ dnorm(-1.096, .25)

# main effects

lamda1[i] ~ dnorm(0, .25) T(0, )

lamda2[i] ~ dnorm(0, .25) T(0, )

lamda3[i] ~ dnorm(0, .25) T(0, )

# interaction effects

lamda12[i] ~ dnorm(0, .25) T(0, )

lamda13[i] ~ dnorm(0, .25) T(0, )

lamda23[i] ~ dnorm(0, .25) T(0, )

}}
```

Appendix B

MCMC analyses

Table B.1 presents the information for all MCMC analyses across all conditions. To reduce the autocorrelation and save the memory, each chain was thinned by 1. For all replications, only last 5000 iterations were kept to provide summary of MCMC analyses.

Table B.1

Information for MCMC Analyses

N	MO	Number of Chains	Chain Length	Burn-in
100	3	2	25000	20000
100	5	3	30000	25000
200	3	3	30000	25000
200	5	4	45000	40000
300	3	4	45000	40000
300	5	5	55000	50000

Note. N represents the sample size; MO represents the number of measurement occasions.

Regarding priors, non-informative priors were used for the LCDM item parameters, growth factor parameters. However, to improve the growth model convergence, the true variance-covariance matrix parameters were used as the prior for the estimated variance-covariance matrix parameters.